NOV 20 1930

...GICAL ROOM
...AL LIBRARY
... OF ILLINOIS

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

HAROLD RUGG, *Editor*

VOL. XXI, No. 8 — NOVEMBER, 1930 — $5.00 per year

CONTENTS

DISTURBERS OF TETRAD DIFFERENCES. SCALES. *C. Spearman* 559

HABITS, INTELLIGENCE, AND ANALYSIS. *William F. Bruce* . . 574

A COMPARISON OF STATISTICAL MEASURES OF OVERLAPPING WITH CHARTS FOR ESTIMATING THE VALUE OF BI-SERIAL *r*. *Percival M. Symonds* 586

STUDIES IN HANDEDNESS: I. A TECHNIQUE FOR TESTING UNIMANUAL HANDEDNESS. *R. H. Ojemann* 597

INTELLIGENCE OF COLONY PEOPLE. *W. W. Ludeman and J. R. McAnelly*. 612

WHEN DOES A TEST MEASURE THE SAME FUNCTIONS AT ALL LEVELS? *Barbara S. Burks*. 616

A FURTHER NOTE ON THE CORRECTIONS OF CERTAIN ERROR FORMULAS. *Harl R. Douglass* 621

NOTE ON THE STANDARD RESPONSE ERROR IN A MEASURE OF IMPROVEMENT. *Jack W. Dunlap and Edward E. Cureton* . 625

CORRECTIONS 626

A REPLY TO PROFESSOR SPEARMAN. *Harold Dean Carter* . . . 627

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION. 629

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

Editor: HAROLD RUGG

Board of Associate Editors

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — E. L. THORNDIKE
FRANK N. FREEMAN — V. A. C. HENMON — L. M. TERMAN — GOODWIN B. WATSON

Assistant to the Editor: FRANCES M. FOSTER

The Journal is published monthly, except June to August, by Warwick and York, Inc., Baltimore, Md., and York, Pa.

The price is $5.00 a year, payable in advance; foreign postage 30 cents extra. Single current issues are 70 cents each, and subscriptions for less than a year are charged at single issue rate for each number ordered. Back volumes are $6.00 each; single back numbers, 90 cents each, except, when more than three years old, the price is $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence concerning editorial matters should be addressed to Harold Rugg, Editor, 425 West 123rd Street, New York City. Subscriptions should be sent to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

TEACHERS' HEADQUARTERS

European

Plan

Entirely

Fire-Proof

Centrally

Located

EDWARD DAVIS

Manager

HOTEL RENNERT

BALTIMORE, MARYLAND

The Psychology of Learning

1928 Revision — BY WILLIAM H. PYLE

$2.20 + 12¢ postage

Warwick & York, Inc., *Baltimore, Maryland*

The Acid Test

What more exacting test can there be of dictionary leadership than the judgment of the educational world?

Every State that has adopted an unabridged dictionary for use in its schools has chosen exclusively

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The Colleges voted overwhelmingly in favor of Webster as the standard of pronunciation, in answer to questions submitted by the Chicago Woman's Club.

Nearly 100% of all schoolbooks are based on the New International for pronunciation, spelling, compounding, and division of words.

The New International has been universally accepted by those best fitted to judge.

It is in accord with the best modern practice.

Use it in your own schools.

Write for our free booklets of interesting ways to teach the use of the dictionary.

GET THE BEST

G. & C. MERRIAM COMPANY
SPRINGFIELD, MASS.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXI | November, 1930 | Number 8

DISTURBERS OF TETRAD DIFFERENCES. SCALES

C. SPEARMAN

THE PROBLEM

In the course of controversial writings about the theory of mental ability, there has often been an urgent demand for correlations and tetrad differences derived from extremely numerous cases. Not uncommonly, at least a thousand have been declared to be indispensable. On one notable occasion, the call was made for no less that ten thousand. The aim of such heroic undertakings has been to bring about some decision between two contending theories. According to one of these, the correlations between different abilities are due to their possessing some general common factor. According to the other theory, they spring instead from the fact of every ability consisting of a random sample of elements; the capacity of every individual is held to vary randomly from one element to another.

Now, the hope of settling this dispute by employing very numerous cases rests on the following grounds. Such cases will, of course, make the "probable errors" very small; and then, considered in comparison with these, even small observed tetrad differences will have the opportunity of showing themselves to possess significance. Should they show this—these writers assume—then the theory of a general common factor must be rejected and the other one adopted instead. But should the event fall out contrariwise and the tetrad differences still remain insignificant despite the shrunken sizes of the probable errors, then and then only these writers concede that the sampling theory must give way to that of a general factor.

But with all this argument the present writer is unable to agree. In general, the two theories are far from standing in any such mortal opposition to one another as is commonly believed. There is nothing to prevent *both* from being true at the same time; or, for the matter of that, neither.

As for the evidence to be obtained from making the cases very numerous, let us first consider the supposition that, on so doing, the tetrad differences still remain insignificant. This does not really deal any fatal blow to the sampling theory; for the latter can always parry the numerousness of the cases by assuming that the hypothetical elements are correspondingly numerous; by this assumption one can make the tetrad differences as small as ever he likes, and thus always maintain his theory in perfect agreement. Truly enough, this assumption plunges the theory into a new and worse trouble; for as the number of elements grows very large, all individuals must necessarily tend to have equal abilities, an inference in flat contradiction with universal experience.[1] But since this new trouble has nothing to do with the tetrad differences, we need not here concern ourselves with it.

Let us now turn to the alternative and contrary supposition; namely, where, by reason of the cases being very numerous, the tetrad differences do become significant in comparison with their probable errors. Would such a result really be, as so commonly believed, destructive to the theory of a general factor? Quite the reverse; such a result might only be just what the theory itself demands. For how great ought, according to this theory, the tetrad differences to be? Under certain ideal conditions, as everybody knows, they should be zero. And under actual conditions, as is also well known, disturbances must be expected to derive from the errors of sampling;[2] these disturbances are the ones that can be eliminated to any degree by making the cases more and more numerous; their general magnitudes are indicated by the so-called probable errors. But I would urge that there also exist *other disturbances*, most of which, far from being removed by increasing the number of cases, are even augmented thereby. In order, then, properly to test the theory of a general factor, the observed tetrad differences should be compared not merely with their probable errors but with the aggregates of *all* the disturbances which by the theory ought to occur. If, and only if, this whole aggre-

---

[1] See: The Sub-structure of the Mind by the Present Author. *British Journal of Psychology*, Vol. XVIII, 1928.

[2] This sampling must not be confused with that which the above-mentioned "theory of sampling" advocates. It is a fact of undenied experience, an actual selection of a sample out of an actually observed population. Whereas the said theory proposes a purely hypothetical sampling out of no less hypothetical elements, the whole being imagined solely for theoretical purposes.

gate of expectable disturbances does not match with those actually observed, then indeed the theory of a general factor would have to undergo at least amendment.

If that were to happen, we should at last really have good ground for trying to obtain fundamental assistance from any rival theory, such as that of sampling. Much work would be needed. The procedure should begin with calculating what amount of tetrad difference remains unaccounted for; that is to say, remains over after making due allowance, as we have seen, for *all* the disturbances for which the explanation is already known. From the size of such a residual tetrad difference, it should become possible to calculate what number of elements the sampling theory ought to assume. Finally, this number should be verified by further observations and considerations; research could show whether repeated determinations of this number of elements proved to be consistent with one another and with all other known related facts. As yet, however, no such serious investigation of the sampling theory appears to have been even contemplated by any of its advocates. And, indeed, the obviously needful preparation for it all has hardly yet begun; that is to say, the ascertainment of the chief disturbances that are demanded by the theory of a general factor, if this theory is to be regarded as *exact*. To contribute toward this preparatory work is the hope of the present paper.

List of Disturbant Influences

Let us, then, pass these disturbances in rapid review, touching chiefly on those points about which misunderstandings appear to be specially prevalent at the present day.

First may be put on the list the trouble of which we have several times spoken already, as being indicated by the probable errors. Such errors are wholly and solely derived from the fact that research cannot deal directly with the entire limitless "population" which is scientifically at issue, but only with a sample of this; they should be called the probable errors of sampling. Of all the disturbances of the tetrad differences, this one alone can be eliminated by making the sample consist of very numerous cases.

Next among the influences which prevent the tetrad differences from being zero may be cited the group factors induced by excessive likeness between the abilities at issue. Here we ought especially to realize that the degree and kind of likeness capable of producing a

group factor never was or could be laid down *a priori*. Instead, it was necessarily left to be decided by actual observation. The theory of a general factor is perfectly satisfied if *any* degree of unlikeness can be found where the group factors disappear and yet some correlation remains.[1]

Distinguishable from these group factors due to excessive likeness are those caused by accidental linkage between the abilities. This point acquires particular importance in the case of school subjects. For example, those pupils who give most study to physical science will naturally give an extra amount of attention to mathematics also; and so the two become artificially correlated together. On this account, the common practice of investigating the theory of the general factor, or *g*, by means of scholastic work or tests must be deprecated.

Next may be mentioned the disturbing influence of what has been called the heterogeneity of the subjects; especially in respect of race, age, sex, education, and so forth. During the last year or two this influence has been much exploited; it has even been made the basis of an attempt to show that the *g* is, after all, spurious. Eliminate the heterogeneity, it has been claimed, and then most, if not all, of the *g* will vanish also. But these recent writers overlook the fact that such heterogeneity and its effects were studied with great care a long time ago; indeed, in the very earliest works on the theory of two factors. The sole novelty in the new view is that now the effect of heterogeneity is depicted as something essentially erroneous; comparable, for instance, with that of sampling errors; or of "attenuation." Whereas, according to the older and, I believe, more correct view, it is only erroneous when and in so far as it happens to be temporarily irrelevant. For example, let us suppose that a population heterogeneous in respect of age is submitted to tests some of which depend on speed of handwriting. Conceivably, the older children might write much faster than the younger ones, although as regards *g* not much superior to them. Here the differences of age are undeniably harmful; for we happen to know that the speed of writing is something very different from *g* and is therefore here an irrelevant and confusing intruder. But now suppose instead that the conditions of the case do not bring into play any such irrelevant factor as that of handwriting; let the tests be correlated with age for no other reason than that both are correlated

[1] Contrary to what has been stated elsewhere, there is no *a priori* necessity that any such degree should exist; quite conceivably, the group factors might not disappear until the correlation did so too.

with *g*. This time the heterogeneity of age may for many purposes, though not for others, be proper enough; it may be quite desirable to take into account not always the lesser differences of *g* that occur when age is constant, but sometimes instead the augmented differences of *g* that occur when age has considerable variation.

All the foregoing influences which prevent the tetrad differences from being zero were specified and emphasized at the very beginning of the whole theory of *g*. Their effects were so gross as to be detectable forthwith. But even in those early days it was expressly anticipated that further sources of disturbance would be discovered later on. The course of measurement in psychology was almost certain to follow the same general lines as in physics. The larger errors are eliminated more and more exactly; while there are simultaneously brought to light further and further errors of continually diminishing size.

And, indeed, something of this sort has actually happened even with respect to the group factors caused by excessive likeness. The degree and kind of likeness that can have this effect have been determined much more precisely than was at first possible. The original observations had only justified such vague conclusions as that the likeness should be a "conspicuous" one. Subsequent observations have for the most part tended to show that the requisite degree of likeness was even greater than at first anticipated; again and again abilities have failed to reveal group factors in spite of appearing very alike indeed. But in some rare instances the reverse has occurred; group factors have arisen from lesser degrees of likeness than had been found effective originally. Most important in this respect have been the results obtained by Davey.[1] This investigator found that some of the verbal tests currently put together in the same team as being different ones are really alike enough to cause group factors of appreciably magnitude. The proof that the disturbance here was not due to any general factor of verbality was furnished by the fact that only some of the verbal tests were thus disturbed, not the others. Furthermore, when two sets of tests were constructed which differed solely in the one being verbal and the other non-verbal, then all signs of the group factors disappeared. It is astonishing that this momentous work of Davey on verbal factors should have been ignored by subsequent writers on the subject.

[1] *British Journal of Psychology*, Vol. XVII, 1926. For a fuller account, see her thesis in the library of the University of London.

Another and very recently discovered kind of group factor is a curious one linking together all the tests in the team constructed by one and the same psychologist. This important observation of idiosyncrasy was made by Stephenson.[1]

Due to the same investigator is the detection of another source of disturbance, deriving this time from propinquity in order of occurrence. If, say, eight successive tests are given, such tetrad differences as $r_{11} \cdot r_{88} - r_{18} \cdot r_{81}$ tend to have positive values irrespective of the nature of these tests in themselves.

More serious still, however, has been found the disturbance introduced by dealing with either the accuracy or the speed of a performance separately. Each by itself is only a partial measurement of the ability. To obtain the total and alone valid measurement of it, both accuracy and speed ought to be suitably combined. Such partial measurements are responsible for much fallacious discussion of a "speed factor."

Another trouble has been noted in the not infrequent practice of testing the subjects—especially when these are very numerous—in separate sections or batches. A historical instance was the treatment of the "special experimental group" in the testing of the American Army. This group was made up of several sections, each tested in a different camp and even by different testers. With such a procedure, group factors seem almost inevitable. Possibly this effect could be minimized by calculating the correlations for each section separately and then averaging.

Yet another disturbance—not indeed so far actually observed, but to be expected *a priori*—derives from the fact that the zero tetrad differences have only been shown to derive from the two factors g and s when all the abilities are simply *linear functions* of these. Whereas almost certainly such simple functions are not the exact truth but only first approximations to this. Akin to this difficulty is that which comes from the fact that measurement—whether physical or mental—always involves the usage of scales. For these can be constructed in different fashions, leading to different results. Only one result can be correct, so that the remainder can be no more than approximations.

Besides all the preceding disturbances which admit of being formally enumerated, there must be an infinite number of others too varied and too irregular for any cataloguing. They may be negligible

---

[1] See his thesis in the library of the University of London.

enough when the cases are few, but not so when these are very numerous. They include a myriad of small details of procedure which may through long habit pass unnoticed and at any rate are not often published. Rarely does the author mention even whether he has made use of Sheppard's correction, or how many decimal places he has taken into account, or how, if at all, his calculations have been checked. There are only too good grounds to suspect that a large proportion of the results that find their way to print are definitely—often even grossly—miscalculated.

Such, then, are the chief influences known already, owing to which the tetrad differences should, by the theory of a general factor, tend to values *not* precisely zero. If these values prescribed by the theory are a good match for those actually observed, then this theory will —so far as ever possible—have shown itself to be exact. In particular, no room will be left to admit the existence of any further influences such as are conceived by the sampling theory.

Variations of Scale

We are left, then, with the task of helping to determine more completely than hitherto the magnitude of the tetrad differences that should be expected from such disturbances as have just been enumerated. And there is one of these in particular with respect to which, as already mentioned, the ensuing tetrad differences have not yet been submitted to estimation at all; it consists in the possibility of variation of scale. To this disturbance we will devote the rest and the most important part of the present paper.

Every measurement that goes beyond mere equality and inequality is faced by the problem of dividing up the extent at issue into comparable parts. For certain purposes and in certain respects—notably those of space, time, and frequency—to do this may be easy enough. Thus a length of ribbon can without difficulty be divided into so many feet, each of which may by juxtaposition be shown to equal each of the others. Accordingly, one might say reasonably enough that the difference between one foot of the ribbon and two feet is exactly the same as that between nine and ten. Indeed, such a view is commonly used as the basis for calculating prices. Nevertheless, even here, in the comparison of spatial extents, some embarrassment may arise. Suppose, for instance, that one man can jump over the height of six feet; another, only three. To state that the former man can

jump just twice as well as the latter one is, to say the least of it, a very inadequate description.

Still worse becomes the problem in the case of measuring by means of mental tests. Often, indeed, marks, scores, or points can be assigned easily enough. But do these always supply legitimately comparable units? Suppose, for instance, that the highest possible mark is 100. Shall we say forthwith that the difference between 0 and 1 mark is just the same as between 50 and 51? Or, again, the same as between 99 and 100? None would venture to make any such assertion. Very often the assignment of the marks is altogether arbitrary. This happens, for instance, when three marks are given for an answer which is perfectly correct, two when it is nearly so, and one when it is far from correct but not completely incorrect. In other ways the marking may be made more different still. Think, for instance, what widely unlike scales are yielded according as a man's memory is measured by the amount he can learn in a certain time, or by the time he takes to learn a certain amount.

As a difficulty in itself and apart from any reference to tetrad differences, this problem of scaling is one that leading psychologists have by no means neglected. To escape from the difficulty, diversity, arbitrariness, and incommensurableness encountered in assigning marks, other radically different ways have been devised of rating ability. Some of these merely alter the unit and the origin of measurement, and therefore would have no effect whatever on correlations. A drastic change, however, has been made by exchanging the absolute marks for ratios; thus a reduction of errors from twenty per cent to ten per cent might be taken as equivalent to one from ten per cent to five per cent because the ratios are the same. But this plan has had more success in economics than in psychology. Still more revolutionary, and this time of wide psychological usage, is the device already conceived and much employed by Galton, where the scale of marks is replaced by one derived from the frequency distribution. That performance which can only be done successfully by ten persons out of a hundred was taken to delimit the highest unit of the whole extent; it was called the top "decile." From this performance to the one that can be done by the next best ten persons out of a hundred constitutes the second decile. And so on, up to the tenth decile, with which, of course, the scale is complete. Obviously, these units can be further subdivided in similar fashion. Or, the decimal system can be exchanged for any other one; thus, very commonly the ten deciles have been replaced by four "quartiles."

Quite different again is the scale introduced by Binet, since this depends essentially on age. The unit consists of the difference

Fig. 1

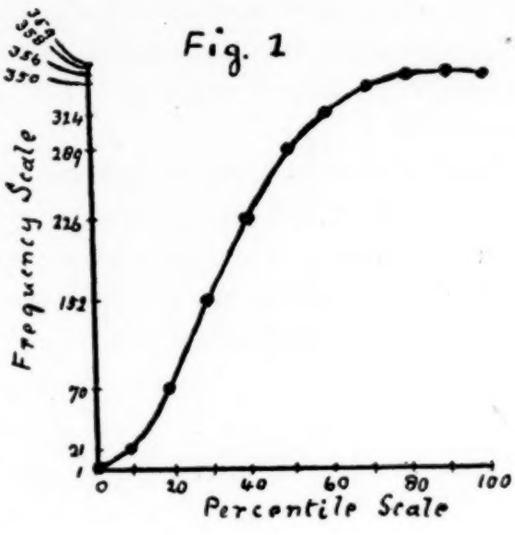


Fig. 4

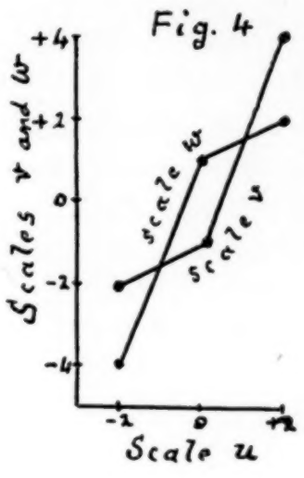


Fig. 2

Fig. 5

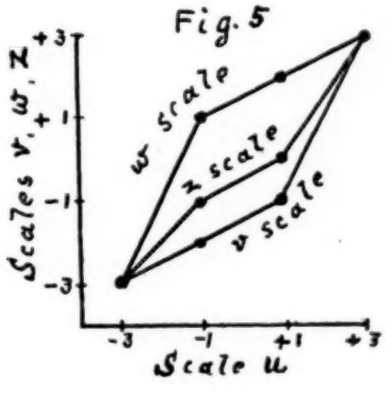


Fig. 3

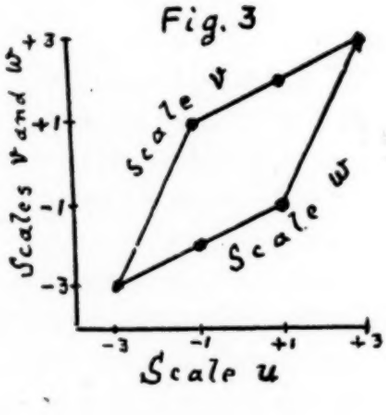
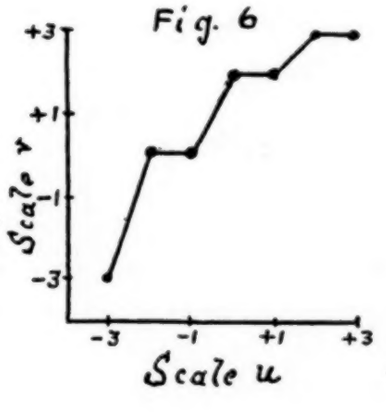


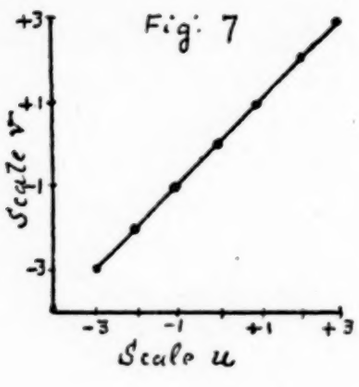
Fig. 6

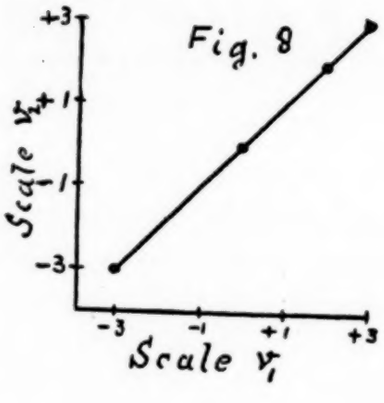


between the performance that some standard per cent of persons can do at any number of years of age and the performance that the

same per cent of persons can manage when they are a year older. Such a scale—the "mental age"—has proved extraordinarily captivating with those who have been more educationists than psychologists. And undeniably, it has achieved a wonderful success in actual practice. Yet at bottom, nothing could theoretically be more unsound. For instance, it implies the absurd corollary that persons go on steadily improving in intelligence all their lives!

Recently a more complex method of constructing scales has gained much repute. It is a combination of the system of marks with that of percentiles. Those differences of marks are taken to be equal which would be so *if* their frequency distributions were "normal." In this way, and this alone, it has been emphatically claimed, the





differences at different parts of the scale are at last truly equal to one another. To this method the present writer is very partial himself. But when he looks about for the definite evidence that it possesses any such eclipsing virtue, he finds it elusive.

Yet further scales have been proposed, some of which may have an important future. A clever one has been suggested by Thurstone, which makes use of age again, but this time in a new and incomparably more scientific manner. Promising work in the same sphere is being done by Courtis and others.

However, my present purpose is by no means to assess the different scales in respect of merit, but only to stress the well-known fact that they do differ from one another, and very widely. This may be exemplified in the following data supplied by Kohs, where the same actual performance (grip with a dynamometer) is shown, in the first place on the percentile scale, and secondly on the scale derived from the hypothetical normal frequency distribution. The following

table gives the corresponding values obtained by the two methods respectively.

Percentile	Frequency	Percentile	Frequency	Percentile	Frequency	Percentile	Frequency
0	14	30	25	60	37	90	48
10	18	40	29	70	40	100	51
20	22	50	33	80	44		

The same state of affairs may be represented graphically by the curve in Fig. 1, where the abscissæ are the scores for gripping, while the ordinates give the frequencies. The ensuing curve, or line of correlation, is evidently far from linear.

Now what, if any, effect can such diversity of scaling have upon the tetrad differences? In order to display such influence of scaling, the first necessity is to clear the ground by eliminating the influence of the errors of sampling. That is to say, the cases must be made infinitely numerous. And as this condition cannot be satisfied naturally, it must needs be created synthetically. And this is easy enough. It can be done by the formula for the correlations of sums and differences.

In this way, our first experiment has consisted in assigning to g two scales having the values, 1, 2, 4 and 1, 3, 4 respectively (each value with equal frequency). Then there are introduced four different specific factors, or s's, each having the values 1 and 3 (equal frequency). Out of all these four variables are constructed: a, which is $g_1 + s_1$; b, which is $g_2 + s_2$; c, which is $g_1 + s_3$; d, which is $g_2 + s_4$.

For the two scales of g the line of correlation is shown in Fig. 2. Again it is markedly non-linear, though perhaps not more so than in the preceding instance. But the correlations that ensued between the four variables are as follows.

	a	*b*	*c*	*d*
a	....	.565	.609	.565
b	.565	....	.565	.609
c	.609	.565	....	.565
d	.565	.609	.565	

There are three different tetrad differences (irrespective of sign) and these have an average value of no less than .034. From this we conclude that the fundamental theorem, whereby variables of the form $g + s$ with no group factors make the tetrad differences zero,

ceases to apply when the g is scaled in different ways. Hence, since the theorem is true without possible exception, we can infer that a mere change of scale suffices to generate group factors.

In the next experiment, g was given the values of $+2$, 0, and -2 (with equal frequency), while each of the four s's again took $+1$ and -1. In consequence, each of the total variables $(g + s)$ took the values $+3$, $+1$, -1, and -3 (the inner values occurring twice as often as the extreme ones). As so constructed, all the tetrad differences must necessarily be zero, since this time g is every time scaled in the same way. But now we will try the effect of scale changes that this time are not made in g separately but in the total variables. Calling the original scale u, we will introduce two others, v and w, constructed as follows.

Scale u	$+3$	$+1$	-1	-3
Scale v	$+3$	$+2$	$+1$	-3
Scale w	$+3$	-1	-2	-3

This table means that $+1$ in scale u changes to $+2$ in scale v and into -1 in scale w. Similarly, -1 changes to $+1$ and to -3. The extreme values make no changes. Expressing the changes as before in correlation lines, we get Fig. 3.

The ensuing correlations between the four variables, each of them put into all three scales, are given below.

	u	v	w
u	.7274	.6365	.6365
v	.6365	.5909	.5228
w	.6365	.5228	.5909

The tetrad differences, arranged in order of magnitude, are: .0247, .0247, .0249, .0422, .0432, .0432, .0671, .0757, with an average of .0412, which is even larger than before.

The next trial, however, showed surprisingly different results. The scale of values $+1$ and -1 (with equal frequencies) was given to g and to each of the four s's. Hence, $g + s$ had each time the scale of values $+2$, 0, and -2 (the zero values occurring twice as frequently as either of the extremes). This scale being called u, two further scales, v and w, were constructed as follows.

Scale u	$+2$	0	-2
Scale v	$+4$	-1	-2
Scale w	$+2$	$+1$	-4

The lines of correlation between u and v as also between u and w are shown in Fig. 4. The non-linearity looks just as great as before. But the correlations turn out to be as follows.

	u	v	w
u	.5000	.4527	.4527
v	.4527	.4527	.4527
w	.4527	.4527	.4100

And from these correlations we find that now—in spite of the considerable changes in them, owing to the changes of scale—*all the tetrad differences remain zero exactly.*

In the next trial the variation was introduced of giving to g four times greater variance, and therefore greater weight, than to s. Thus, g had the values of $+2$ and -2, whereas each s only had, as usual, $+1$ and -1. In this way $g + s$ always had the scale of values $+3$, $+1$, -1, and -3 (each with equal frequency). Out of this scale u were fashioned the following scales v, w, and z.

Scale u	$+3$	$+1$	-1	-3
Scale v	$+3$	-1	-2	-3
Scale w	$+3$	$+2$	$+1$	-3
Scale z	$+3$	$+2$	-2	-3

The correlation lines of u with v, w, and z are shown in Fig. 5, and they look quite as diverse as any we have seen. Extraordinarily unequal are, too, the correlations, given below.

	u	v	w	z
u	.800000	.687208	.687208	.877253
v	.687208	.590361	.590361	.753430
w	.687208	.590361	.590361	.753430
z	.877253	.753430	.753430	.961538

Yet despite all this, every tetrad difference is again *exactly* zero (that is to say, of course, exact when worked out in an exact manner; not, as here, using only six places of decimals).

The results of the two preceding trials were at first extremely puzzling. But eventually I attributed them to the fact that changes in the tetrad differences derive only from the scaling changes of g, not from those of the s's. But in both the two preceding trials, g was originally allowed only two values; whereas for a change of scale to be possible, the values must originally be at least three in number. Anyway, these two instances suggest that the handling of this whole matter requires no little circumspection.

The aim of the next and last trial was to see whether any tetrad differences other than zero make their appearance when four variables on the same scale are transferred to another one, but still the same for all four alike. For this purpose, g was given the values of $+1$, 0, and -1, while g_2 had those of $+2$, 0, and -2 (with equal frequency in each case); s_1 and s_2 were made like g_1, while s_3 and s_4 were like g. Out of all these components the following variables were constructed: $g_1 + s_3$, $g_1 + s_4$, $g_2 + s_1$, and $g_2 + s_2$. Each of these variables, obviously, has just the same total scale, which we will call u. This was then for all four variables transformed into v.

Scale u	+3	+2	+1	0	−1	−2	−3
Scale v	+3	+3	+2	+2	0	0	−3

The correlation line between u and v is as shown in Fig. 6, and is naturally far from linear. Whereas the line for any two different variables both in scale u is linear exactly, as shown in Fig. 7. The same is no less true about the scale v. (See Fig. 8.) Now turn to the correlations. For the u scaling they are as follows.

	u_1	u_2
u_3	.200	.400
u_4	.400	.800

That is to say, the tetrad difference is exactly zero, as one would expect. But now take the correlations for the v scaling.

	v_1	v_2
v_3	.267	.400
v_4	.400	.689

In spite of the perfectly linear form of the correlation line, the tetrad difference comes to as much as .0237.

Reviewing all these results obtained from synthetically produced variables, the alteration of scale effected seems quite comparable with what occurred in the actual instance given by Kohs. If that be so, the tetrad differences found by us are at least quite within the range of possibility, though probably a good deal larger than what occurs very commonly. Most usual, perhaps, are amounts between .01 and .02.

To apply this finding in a general way is easy enough. We may, pending more accurate information, conclude that the tetrad differences should, on the theory of two factors, not range about the p.e. as commonly believed, but instead about $\sqrt{\text{p.e.}^2 + .01^2}$ or $\sqrt{\text{p.e.}^2 + .02^2}$.

This amendment will make little difference so long as the cases are few; but it may become most important when the cases are very numerous.

Any more specific application of our results would appear to be a much harder task. It seems natural to expect that the influence of change of scaling will reach its greatest degrees where the frequency distributions of the variables depart most widely from the "normal" type. And one is even tempted to hope that much of this influence may be practically eliminated by rescaling the variables so as to render the frequency distributions normal. Possibly, too, some benefit may be derived from replacing the correlation coefficients by correlation ratios. But before anything of this sort can be done with legitimate confidence, probably much more investigation of the matter will be needful.

In conclusion, it may be added that the preceding inquiry would seem to have produced incidental results of interest for quite other topics than that which are specially in view. The most surprising of these has been the enormous fluctuations which changes of scale have produced in some of the correlations, even when the saturations of the variables with g were left untouched. In one table, for instance, the correlations varied from .59 to .96! If the former value had in actual practice occurred between two tests of general ability, these tests would be considered very bad; whereas the latter value would be taken to prove them extraordinarily good. And yet in the two cases the saturation of the tests with g—and therefore their real excellence—remains just the same. Here is a very novel and yet fundamental consideration both for the construction of mental tests and for their eventual appraisement.

HABITS, INTELLIGENCE, AND ANALYSIS

WILLIAM F. BRUCE

State Normal School, Oneonta, N. Y.

According to an ancient notion, habitual acts should be distinguished sharply from intelligent activities. This idea still persists in education, and today it leads frequently to an overemphasis upon habits. Methods center in habit formation, and curricula are built in terms of habits to be established. This current trend began a generation ago when psychologists turned away from the introspection of mental faculties and devoted their energies to the experimental examination of physiological changes occurring during learning—chiefly to the study of bodily habits. William James also gave a practical turn to teachers' attitudes toward habit by eradicating the popular notion that habits are usually bad habits.[1] Thorndike has followed James in stressing the inculcation of *useful* habits, but has gone a step further by declaring, "The *one thing* that [schools and other educational forces] can do best is to establish those particular connections with ideas which we call knowledge and those particular connections with acts which we call habits."[2] Although knowledge is mentioned separately here, Thorndike has maintained that the acquisition of all knowledge is included under the "one thing" which is, for him, habit formation. Consequently, writers on education now commonly make habit central. We read, for example, that a psychology of elementary school subjects "deals with the formation of habits that give skill in their acquirement and all the factors that influence those habits."[3] Or, that "Every teacher knows nowadays that the business of education is the business of forming habits."[4] Such statements give the impression that habits, as specialized parts of the organized life, are more important than the unity of organization which is represented by the term intelligence.

The contrast which is implied between habits and intelligence raises a problem so difficult that the instructor in educational psychology is tempted to avoid the issue when a student asks him to explain the relation of habit to intelligence. One way of side-stepping involves

[1] *Cf.* James, William: "Talks to Teachers." 1899, p. 64.

[2] Thorndike, E. L.: "Educational Psychology, Briefer Course." 1915, p. 401.

[3] Reed, Homer: "Psychology of Elementary School Subjects." 1927, p. 1.

[4] Stillman, Bessie: "Training Children to Study." 1928, p. 238.

the substitution of "intelligence" for "personality" in Watson's statement, "Our personality [intelligence] is but the outgrowth of the habits we form."[1] In this answer intelligence disappears as though it had been absorbed by habits. Another evasion faces in the opposite direction as it quotes the old analogy, "Habits are the tools of intelligence." Here intelligence is granted a verbal dominance. A third device draws a sharp line between the two kinds of activity by presenting intelligence as an analytic process in contrast with habits, which are labeled "unanalytic."[2] This last treatment separates habits from intelligence without indicating any method of comprehending their relationship. These three replies are typical of the way in which educational psychology fails to face the fundamental issue of resolving the dualistic conflict exemplified in these two terms. Meanwhile, antiquated distinctions between mind and body, which still run strongly in popular thought and common speech, bind students to the notion that intelligence is "mental" and habits are "physical." But a clear view of the educative process can scarcely be attained until misleading analogies have been discarded, overemphasis upon habit has been eliminated, and sharp distinctions that maintain a fallacious dualism have been seriously modified. The problem of habit versus intelligence calls for restatement, and for a more adequate answer derived from the study of concrete behavior.

A brief consideration of the "tool" analogy will show its limitations. While a teacher concentrates on establishing habitual reactions in his pupils, he may salve his conscience by reiterating, "Habits are the tools of intelligence." But is it safe to assume that intelligence will keep the upper hand under this treatment? Are habits really like lifeless tools in the hands of a vigorous workman—"intelligence"? On the contrary, our habits are alive, active, dynamic. They go into action without waiting for the command of an overseeing intelligence. Even when habits have been started through an intelligent suggestion, they often run far beyond the restrictions which intelligence might well impose. This dynamic characteristic of habits has been stressed by Dewey.[3] As he points out, readiness in habits implies "intense craving"—a constant anxiety to jump into action. Habits are not like tools lying on a bench, for they are "energetic and dominating

[1] Watson, J. B.: "Behaviorism." 1924, p. 216.

[2] Judd, C. H.: "Psychology of Secondary Education." 1927, p. 301.

[3] *Cf.* Dewey, John: "Democracy and Education." 1916, p. 57; "Human Nature and Conduct." 1922, pp. 24ff.

ways of acting." Even if one insists on conceiving of intelligence as a distinct process, he cannot fail to see that intelligence must struggle strenuously with dynamic habits if it is to maintain its supremacy. The "tool" analogy is misleading, for it obscures the fact that intelligence fails so often in the control of habits. An analogy that does not clarify the student's view of a particular relation is useless. This one is so vague and lopsided that it adds to the confusion, for the analogy breaks down whenever the student attempts to compare any specific habit with a tool. Perhaps the error lies in attempting to set forth a kind of contrast between habits and intelligence that does not exist. If intelligence is dynamic, so is every habit.

Professor Judd recognizes this dynamic, active characteristic of habits, but he draws a distinction between intelligence and habit based on the assumption that there is a difference in the *kind* of activity. In a section entitled "Intelligence, Analytical; Habit, Unanalytical" Judd maintains that habits "are lacking in analysis and, for this reason, stand in sharp contrast with reasoning processes. Habits are often so compactly organized that the individual who possesses them can give absolutely no verbal description of what he does when he performs habitual acts."[1] An illustration follows which depicts the skillful sawyer who "makes his movements without stopping to analyze the experience; yet he is instantly responsive to the sensations which come from his saw."[2] Thus, an analytic quality which is denied to habit is conferred upon intelligence.

In opposition to the statements just quoted, the writer believes that a contrast between habitual acts and intelligent ones, which is based upon the presence or absence of analysis, is not supported by the facts. In establishing this adverse criticism we must show, first, that the acquisition and maintenance of habitual skill involves analysis; second, that the kind of analysis found in intelligent action is not distinctly different from that which can be discovered in skills; and, finally, that intelligent conceptions are also "compactly organized" to a degree that defies verbal description. The aim of our criticism is to bring habitual and intelligent action within one clarifying conception of the learning process.

Let us begin with a definition of analysis by which we may judge whether or not certain processes are analytical. One dictionary states that analysis is "the examination of the component parts in their

[1] Judd, C. H.: *Op. cit.*, p. 301.
[2] *Ibid.*: P. 30 2

relation to the whole." This definition may be employed tentatively until its use in our discussion reveals more clearly the nature of analysis in human action.

Does not analysis occur during the acquisition of the sawyer's skill? When a novice tries to use a saw—say, in making a straight cut across a board following a line—he has an aim which represents, in prospect, the whole act, from the selection of the saw to the last stroke that cuts off the board. But since the novice lacks intimate acquaintance with many phases involved in the skill, he is quite likely to encounter distinct difficulty as soon as he begins work. Perhaps he has trouble in getting the cut started just on the line. His attention is turned to this specific phase of the task because the result produced at first is not in accord with the general outcome desired. The rectifying of this error involves an examination of one component part of the act in relation to the whole. The learner *is analyzing* the process. As the amateur sawyer proceeds he pays attention to one phase after another because difficulties arise, each of which necessitates variation from the method first tried. Repeated but varied trials aid in discovering new movements which fit appropriately into the whole act. Thus efficient coordination is achieved. Through analytic-synthetic reorganizations a stage in skill is attained when the whole job can be done without "stopping to analyze" its various phases. Suitable adjustments are then made readily. But it is evident that the "compactly organized" habit has arisen out of a series of attempts which have involved many an analytic act.

However, Professor Judd insists that habits are "unanalytical," because "often the skilled mechanic can give absolutely no verbal description of what he does." Analysis is apparently identified with the use of language. But if analysis is to be restricted in meaning to verbal activity, this definition must be justified upon the basis of a distinction which is peculiar to some or to all analyses that employ language. Judd claims that such a distinction exists.

> Language as a medium of thought is distinctly analytical; it makes possible the dissection of situations and their description in most minute detail.
>
> All the relations involved in a train of reasoning are clearly distinguished and brought to vivid consciousness.[1]

The issue comes down to this: Can a line be drawn between the kind of analysis shown in a skillful act and the kind involved in a "train of reasoning" which uses language as its tool?

[1] *Ibid.:* P. 301.

In the first place, note the exaggeration in the statement that often a skilled workman can give "absolutely no verbal description of what he does when he performs habitual acts." If a mechanic uses language at all, some words (in addition to expletives) will be employed, in either overt or implicit speech, whenever an outstanding difficulty is encountered, and as attention is turned to the retrial necessary in that particular phase. Consequently, several points of difficulty will stand out so definitely that the skilled workman can indicate to the novice some phases that require special care. Any language-using human being can give a crude verbal description of difficulties encountered in acquiring a skill, and of methods used in overcoming them. Instead of accepting the misleading overstatement that "absolutely no verbal description" of hand skills can be given, it becomes clear that a crude analytic account is possible in typical cases.

Nevertheless, we must agree that the handworker can never describe his skillful act in "most minute detail." Many a knack is acquired by trying again and again, without the learner differentiating, in words, the exact details of the movement. He *feels* his way through the act rather than *talking* his way into the coordination that occurs. In such cases the words and verbal thinking are sketchy and incomplete. The worker is content when he can *do* the thing, and cares not a whit about *describing* the doing. Even if he does wish to describe the method, as an aid to his apprentice, it is unlikely that he will be able to set forth more than a few of the essential conditions. He is "instantly responsive" in a direct way which brooks no artificial medium like language. The skilled worker is usually much closer to his work than words can take him. He recognizes the meaning of "sensations which come from his saw," but he does not and cannot stop to name each one. Consequently, there is no likelihood that we will secure a detailed description of a compact, habitual process from a handicraftsman, although he has overcome many a minor difficulty, and though he continually makes many fine adjustments by paying attention to this or that specific aspect in an analytic way.

There is a further reason why the worker cannot describe his skillful act, even if he commands the richest vocabulary. A skill is a fluid, adjustive action, ever changing to meet the peculiarities of new situations. Even if an accurate description could follow an adjustive act through the process as it occurred, this description would be quite inadequate for the next occasion, which would differ from the preceding one. A skill is always creative in the sense that the

adjustment is made *as* it occurs, and is not a form of action which has been acquired already in its integrated and adaptive fullness. This adjustability, which is characteristic of habitual skill, defies analysis. Not only is the integration of adjustive skill too smoothly "compacted" for words to measure its exact contour and ramifications, but, in addition, the flexibility of such a habit is too fluid to be caught in a network of words.

Although we recognize the severe limitations of verbal analysis in the realm of habits, such as hand skills, this is not equivalent to admitting that the application of our useful language tool in "intelligent" activities enables one to escape from these limitations. For, when we examine the process of acquiring an intellectual concept, we see how similar are the limitations of verbal analysis in concept formation to those noted in the learning of hand skills. As an illustration, take a student who is trying to grasp the meaning of "instinct" as used in educational psychology. During a prolonged consideration consisting of discussion, reading, and observation, he evolves a concept represented by the verbal symbol "instinct," just as a period of practice with a saw leads to the corresponding skill. It is obvious that during this thoughtful consideration specific difficulties will arise which require special attention in relation to the whole concept. In other words, concept formation involves analysis of the kind defined in connection with the acquisition of skill. Of course, in such concept formation the analysis is distinctly verbal. But does this predominance of words, as indispensable tools in the process, necessarily imply that the resultant—the concept—is not as "compactly organized" as is a skillful act? Can the student dissect this concept "instinct" and describe it in most minute detail?

As far as compactness is concerned, grasping a meaning is very much like catching on to a knack in hand skill. Just as the appropriate twist is discovered in the midst of a trial stroke, so while the student is absorbed in thoughtful consideration an idea "pops into his head." Indeed, since a whole vocabulary of verbal tools is ever ready, the idea that "pops" may not seem very closely related to the thinking which immediately preceded its opportune emergence. This sudden, unpredictable, compact transformation in experience has been noted by numerous writers. Driesch's phrase, "a meaning just comes," may be placed beside Dewey's "leap" in thinking.[1] Bode points

[1] Dewey, John: "How We Think." 1910, p. 75.

to the arrival of a "new stimulus" which is "adequate to the situation,"[1] while Köhler emphasizes the suddenness of insight.[2] These phrases represent different approaches to the central fact that during intelligent activity instantaneous transformations occur, each of which is characterized by a unity or "compactness" which marks the resultant off from the more or less logical steps that form a portion of the experiential background out of which it suddenly emerges.

Considering a wide range of typical cases, we find that at the close of a careful verbal analysis, which precedes the coming of a new conception, there appears in the *dénouement* something more than a mere movement to an inevitable next step. The entire proposition undergoes reformation, all the phases readjusting themselves to each other simultaneously, and thereby being changed in some degree as they become members of the new whole. The uniqueness of the emerging idea corresponds to the complete readjustment involved in a skillful stroke. As the student of instinct redefines his term he makes a "jump" which defies analysis quite as truly as does a habit of the hand.

Although an analysis may be made of steps which have led to the formation of a concept, the incompleteness and inadequacy of such analysis are recognized in the tests employed for the validity of an inference. The final test of inferential thinking and of concepts does *not* lie in tracing backward analytically the logical steps which have led to the conclusion. Thinking is tested by going forward to the application of the conclusion in some concrete situation. Even in the field of mechanics the inventor must demonstrate the accuracy and adequacy of his calculations by an actual trial of his new machine. In other fields, such as biology, psychology, and practical ethics, exact verbal analysis seems even less effective in exposing for critical examination the innermost recesses of concept formation.[3] Analysis may be an aid to thought when it leads to the discovery of a new implication, but this is a movement forward rather than a proving by the retracing of steps.

No matter how thoroughly a student has analyzed any situation, he must always remember that he has reached his conclusion only by jumping to it. Consequently, the new concept must stand trial as a unique *whole* rather than be accepted or rejected on the basis of a

[1] Bode, B. H.: "Creative Intelligence." 1917, p. 248.

[2] Köhler, W.: "Mentality of Apes." 1927, p. 17.

[3] *Cf.* Jennings, H. S.: "The Biological Basis of Human Nature." 1930, p. 372.

wordy dissection. So, habitual and intelligent acts cannot be separated on this score, for while both involve analysis, both are also characterized by an integration which verbal analysis cannot penetrate.

Since an intelligent act, even when employing dry words, is still a human activity, it receives guidance, as does skill, through feeling. As the thinker leaps to grasp his conception, his feelings get their chance to push this way or that. Prejudices, desires, sympathies play their part in molding conclusions. And though reflection may safeguard thinking in some measure by requiring reexamination of the whole proposition, eventually the decision must be reached and deliberation on that particular point must cease. Final judgment that the process of analytic examination has been carried far enough to give a fitting conclusion rests upon the total feeling of the thinker, and this fact may well be compared to the statement that a "feeling of fitness" guides in the acquisition of skill.[1] It is true that through much reflection and many practical trials this fundamental feeling may attain a quality which we call "good judgment." However, good judgment is a unity which may be felt but which cannot be completely analyzed, although the resultant is discovered and expressed by means of words. A felt integration occurs both in skill and in thought, and even though one may long for a higher degree of certainty, he is compelled to trust his own feeling at last. This fact does not imply that a feeling of assurance is the ultimate test in thinking, but rather places emphasis on the necessity of creating a scientific attitude—a feeling of joy while dwelling on the doubtful—to replace that common human desire for certainty which leads to premature judgment.[2] Since we cannot escape from the flexibility of feeling into a world of solid, logical molecules, let us do what we can to make our feelings become helpful guides in thought as well as in skill.

In the light of the preceding discussion it is scarcely necessary to point out that an intellectual concept has a degree of fluidity like that of a skillful knack. Although words preserve meanings, the meaning of a word is not necessarily static, any more than is the pattern of a physical habit. Like habits, concepts range from the relatively inflexible, like "rectangle," to the dynamic, growing meanings, such as "instinct," "evolution," "democracy." In so far as concepts are modified through use, they become too deeply immersed in

[1] *Cf.* Ogden, R. M.: "Psychology and Education." 1926, p. 142.

[2] *Cf.* Dewey, John: "The Quest of Certainty." 1929, pp. 227f.

a moving intellectual world to permit their examination "in most minute detail." Newly created concepts require the invention of new terms or the conferring of new meanings on old names, for each concept has a unique unity that cannot be described fully by a mere summation of other terms. Definitions offer only crude intimations concerning vital meanings, just as descriptions of skills only follow these activities at a respectful distance.

A practical suggestion for teachers arises out of the limitations of analysis in the realm of concepts. Since a concept cannot be completely analyzed, it must be learned by a method similar to that employed in acquiring a skill. The instructor in either case can only point out some of the essential features which seem likely to aid the learner in grasping the knack or the meaning. And these suggested features should vary with the time, the place, the circumstances, and the student. But when the situation has been prepared by the teacher, the learner must make the leap for himself. If he lands awkwardly, the teacher may point out other aspects which may be considered, and thus aid the student in a second trial. The point is that the instructor must always remember that each student is creating a concept of his own with its peculiarly individual tones and colors, just as each skilled workman performs the "same" task in a slightly different way. In teaching we cannot rely solely on analysis of concepts into words, but we must trust the learner to perform acts of an integrative character, the pattern of which cannot be designed in advance, even by the most minute analysis in language terms. So, while we may agree that "language as a medium of thought is distinctly analytical," we cannot agree that the use of this tool of tools makes valid a sharp distinction between learning experiences which involve the widest and fullest use of language and those in which language appears to give little or no aid. Such a distinction between "habitual" acts and "intelligent" ones is misleading, for it engenders overconfidence in mere words.

But it is time to turn toward constructive criticism. In the light of the dynamic and analytic characteristics of habits, and the limitations of verbal analysis, what valid distinctions remain in the contrasting terms habit and intelligence? Can the functioning of language in intelligent activity be disregarded?

Historically, the emergence of language in the midst of a society of tool-users marked an enormous step in social evolution by which the precursor of man increased his effective intelligence enough to justify

calling him a human being. Out of the feeble beginnings of language came foresight, planning, deliberation, reflective thinking. Today each child must acquire language and thereby begin to think. But this step in development should not be confused with our problem. We are concerned rather with the relation between habits and intelligence *after* the learner has passed over into the language-using stage. Language is not disregarded, but it is assumed that this tool is available and ready for instant use, whether the learner is engaged in acquiring manual habits or in delving into intellectual problems.

When applied to concrete acts, the word habit indicates a specialization that has arisen out of the complex of ill-defined activities with which the infant begins life. The hand that at first fumbles vaguely and grasps reflexly eventually acquires specialized skill in holding and guiding a saw, or in the manipulation of typewriter keys. In contrast, intelligence refers to the generalized guidance of activity. An action is commended as intelligent when a general sizing-up of the whole situation has occurred, whereas an act may be designated as skillful and habitual without consideration of its total effects. But it must be remembered that specialized habit and general intelligence grow up together. We cannot accept Watson's view that personality (intelligence) is "*but* the outgrowth" of habits, for this implies that the part precedes the whole. A habit cannot develop except through coordination and integration with other activities which make up the already going concern. Moreover, some degree of coordination and integration is necessary from the very beginning of life. As each habit emerges it must make its peace with the whole life; that is, enter into cooperation with the various conditions within the organism as these are related to the environment. When integration and adjustability reach a high level we label action "intelligent." So there should be no mystery about the relation of these two aspects of life. Specialization and generalized coordination are complementary to each other. A pseudo-problem has been artificially concocted by overemphasizing distinctions, neglecting relations, and forgetting that differentiation and integration progress together in living organisms—even in those that use language.

Another distinction, which is closely connected with the preceding, emphasizes the effect of previous experience upon an habitual action, whereas an act is called intelligent in order to stress the new, creative reorganization which is occurring due to the peculiarities of the present situation, and in consideration of the probable future consequences.

This contrast is also valid so long as the student does not forget that habits enter into the intelligent reorganization, thus making it possible, and that at the same time these habits are reconstructed and become more flexible. As Dewey says, "What is necessary is that habits be formed which are more intelligent."[1] So, when we examine a particular concrete act we find not only that it is in some degree the product of past experience, and therefore "habitual," but also that modifications and adjustments have occurred which make it entirely proper to label the act "intelligent" as well. Any novel action that is marked by a generalized consideration of the present situation and the consequences which may emerge in the future will be found on examination to have roots reaching deeply into past experience and to involve specialization of some sort. In concrete acts the past, the present, and the future come into intimate relation, and likewise specialization and generalized integration occur simultaneously. The adjectives *habitual* and *intelligent* merely indicate different phases to be noted, or a predominance in emphasis by the observer, or the relative location of the act on a scale which includes many gradations from the largely habitual to the highly intelligent.

Our discussion up to this point has dealt with *concrete* acts which for purposes of emphasis may be called either habitual or intelligent. But these terms are also applied to *abstract* characteristics. This "abstract" use becomes a source of confusion unless the student clearly recognizes the difference as he reads psychological literature. For example, Dewey says, "Habit *as such* is too definitely adapted to an environment to survey and analyze it."[2] Is this equivalent to saying that a concrete habit cannot involve analysis? By no means. The term habit as used in this quotation stands for the abstract characteristic of definite adaptation. It is used in this restricted sense as a basis upon which to indicate that something more than definite adaptation must and does occur in effective living. "Impulse" must be released, when definite adaptation fails, and finally "conscious search" or "intelligence" (in the abstract sense) must supervene. Here sharp distinctions are drawn. Each of the three terms is strictly defined and set in opposition to the other two. Such an abstract logical analysis of human action may aid the student in noting how novel suggestions and careful deliberation must occur when old forms of activity come into conflict or prove inadequate for the situation. But the student

[1] Dewey, John: "Human Nature and Conduct." P. 128.
[2] Dewey, John: "Human Nature and Conduct." P. 179.

must remember that many concrete habits involve adaptability, that habits enter into the constitution of our actual impulses, and that in reality intelligent activity does not result in the abandonment of habits. *Habit as such* is unanalytical, but we must acquire *habits* which are more analytical and more intelligent. Abstract statements concerning habit should be so safeguarded by careful examination of concrete habits that no student will be misled into applying an abstract characteristic to a concrete action in such a way as to blind him to its actual flexibility and intelligence.

In conclusion, we may agree that from the standpoint of abstract characteristics it is proper to say that intelligence supervenes following the conflict of habits and the releasing of impulse. But this should not lead to the notion that two separable kinds of activity exist in the same organism, the one being dominant and the other subservient. When a psychologist talks as though concrete habits are inflexible, unprogressive, and unanalytic, while intelligence has contrasting characteristics, he encourages a dualistic view of behavior—a hang-over from the mind-body conflict. And this confuses the student of education. It is less misleading to say that a person's lack of progressiveness lies quite as much in his general intelligence as in his specific habits. Habits are not the tools of intelligence, nor is intelligence the outgrowth of previously established habits. The life of a changing organism in a changing world involves the continual reorganization of the specific phases as well as of the unified totality. Habits and intelligence together contribute to a more abundant life.

A COMPARISON OF STATISTICAL MEASURES OF OVERLAPPING WITH CHARTS FOR ESTIMATING THE VALUE OF BI-SERIAL r

PERCIVAL M. SYMONDS

Teachers College, Columbia University

There are three methods for describing the overlapping of distributions, or for describing the differences between the means of two groups. These methods are: (1) The per cent of one distribution reaching or exceeding the median score of another distribution; (2) the difference between the means of the two distributions; (3) bi-serial r. Each of these three measures also has its measure of reliability. The reliability of the per cent of one distribution reaching or exceeding the median score of another distribution is

$$\sqrt{\frac{.25}{N_1} + \frac{pq}{N_2}}$$

The reliability of the difference between two means is

$$\sigma_{diff.} = \sqrt{\sigma^2_{M_1} + \sigma^2_{M_2} - 2r_{12}\sigma_{M_1}\sigma_{M_2}}$$

The reliability of bi-serial r is

$$\sigma_r = \frac{\left(\frac{\sqrt{pq}}{z} - r^2\right)}{\sqrt{N}}$$

where

$$r = \frac{M_2 - M_1}{\sigma} \cdot \frac{pq}{z}$$

The purpose of this paper is to compare these three methods of describing overlapping. Table I gives comparable values for every .05 of bi-serial r in terms of differences in means and proportion of one curve above the mean of the other curve. In this table it is assumed that the population of the two groups is always equal.

Fig. 1 shows the comparative ratio of each measure to the reliability of the measure. Curve 1 is the ratio of the per cent of one curve exceeding the means of the other to the reliability of this measure.

$$\frac{.50 - p}{\sqrt{\frac{.25}{N} + \frac{pq}{N}}} = \frac{.50 - p}{\sqrt{.25 + pq}} \cdot \sqrt{N}$$

2√N
1·7√N
curve 1 - percent of overlapping.
curve 2 - difference between means
curve 3 - bi-serial r
3
2
1
0 05 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
scale of bi-serial r

FIG. 1.—Ratio of measure of overlapping of two distributions to the standard error of the measure.

TABLE I

Bi-serial r	Difference in means in terms of the standard deviation of one of the distributions	Proportion of one curve above the mean of the other curve
.00	.0000	.5000–.5000
.05	.0799	.5318–.4682
.10	.1601	.5636–.4364
.15	.2411	.5952–.4048
.20	.3233	.6267–.3733
.25	.4071	.6579–.3421
.30	.4931	.6890–.3110
.35	.5817	.7196–.2804
.40	.6735	.7497–.2503
.45	.7694	.7792–.2208
.50	.8701	.8078–.1922
.55	.9767	.8356–.1644
.60	1.0906	.8622–.1378
.65	1.2132	.8874–.1126
.70	1.3467	.9109–.0891
.75	1.4938	.9324–.0676
.80	1.6584	.9513–.0487
.85	1.8458	.9668–.0332
.90	2.0637	.9804–.0196
.95	2.3242	.9899–.0101
1.00	∞	1.0000–.0000

Curve 2 is the ratio of the difference between means to the reliability of this difference.

$$\frac{M_2 - M_1}{\sqrt{\frac{\sigma_1^2}{N} + \frac{\sigma_2^2}{N}}} = \frac{M_2 - M_1}{\sqrt{2}} \cdot \sqrt{N}, \text{ assuming } \sigma_1 = \sigma_2 = 1$$

Curve 3 is the ratio of bi-serial r to its reliability.

$$\frac{r}{\sigma_r} = \frac{r}{\frac{\sqrt{pq}}{z} - r^2} \cdot \sqrt{N}$$

Fig. 2 shows the overlapping of two normal distributions for all cases of bi-serial r from .00 to .95. It may be possible to use this chart to estimate roughly by inspection the bi-serial r, or the amount of overlapping or the difference in means for any two distributions plotted on the same graph.

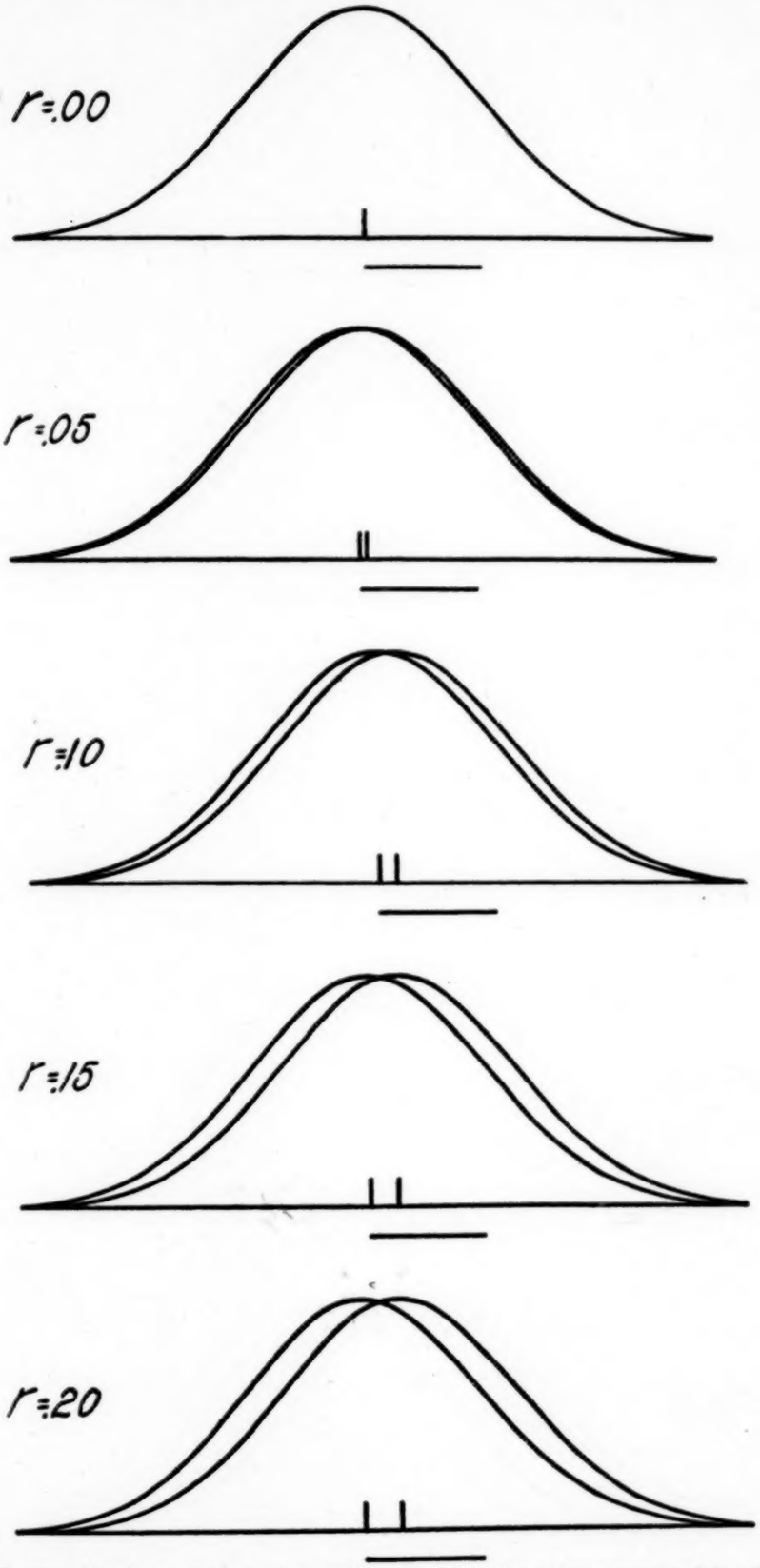


FIG. 2.—Charts showing overlapping of two distributions for every value of bi-serial *r* from .00 to .95.

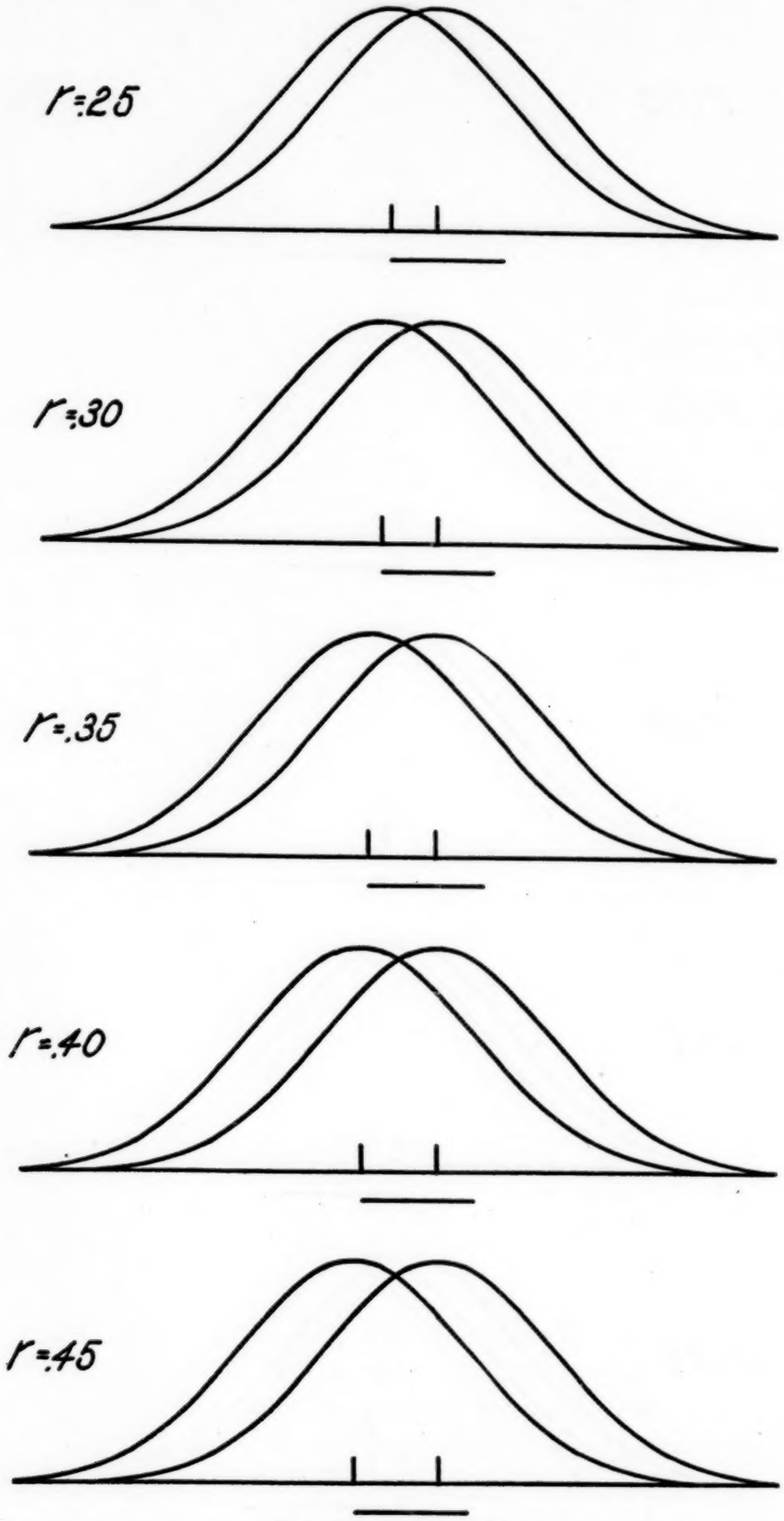


FIG. 2.—(*Continued.*)

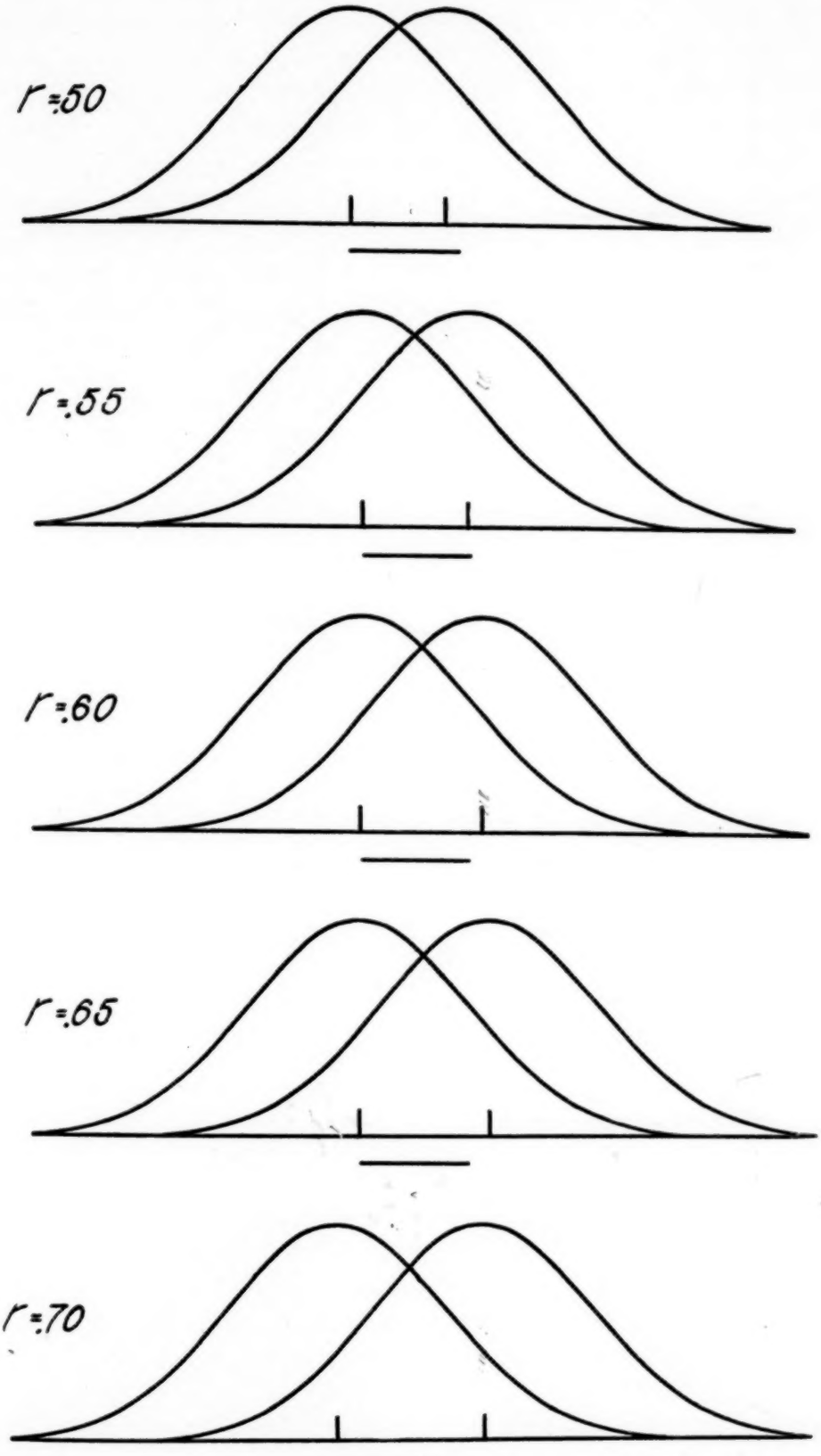


FIG. 2.—(*Continued.*)

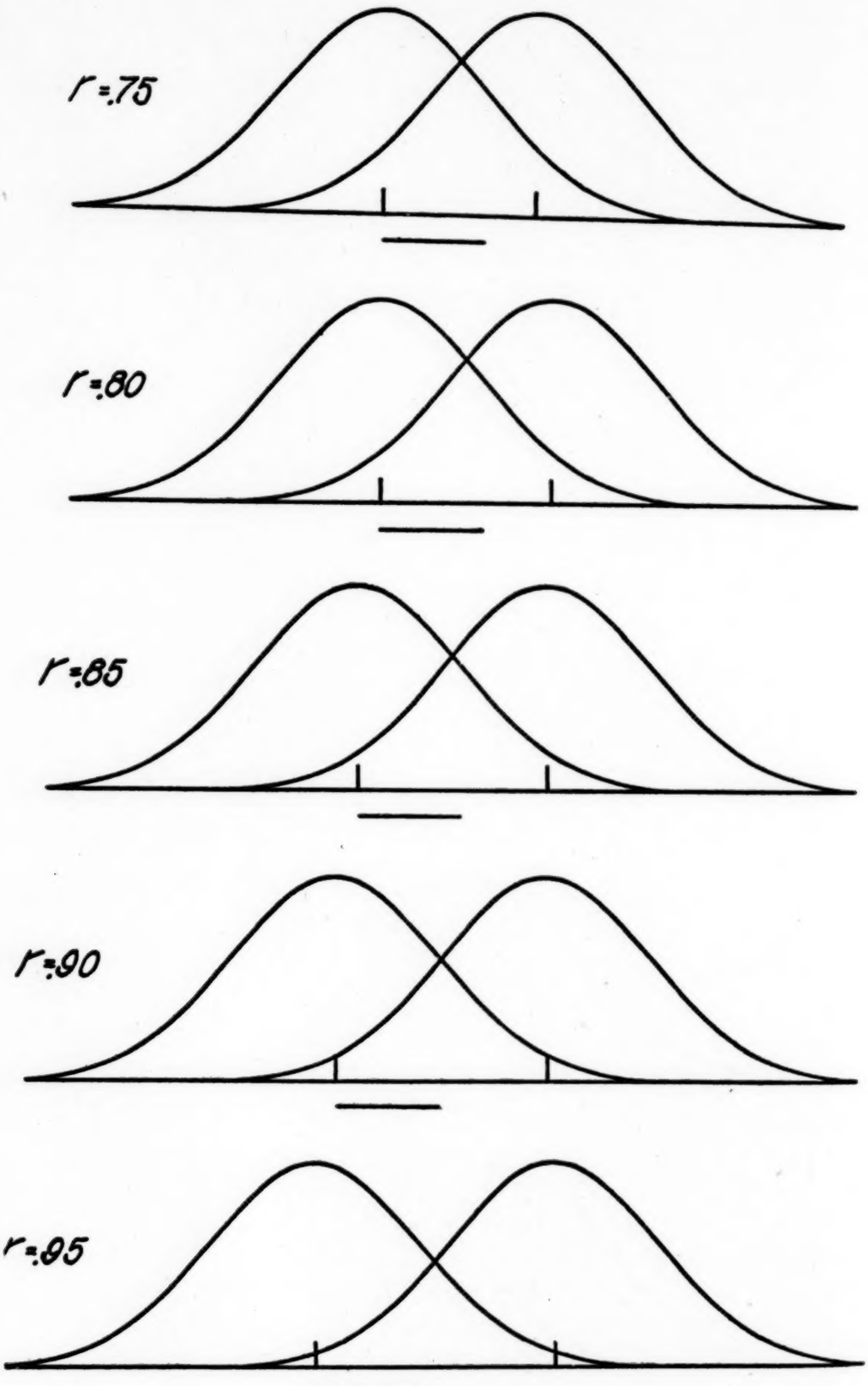


FIG. 2.—(*Continued.*)

The significance of bi-serial r for determining the chances that a given value on the scale of measurement lies in one distribution rather than the other of two overlapping distributions is shown in Tables II and III. Table II shows the *odds* and Table III the *probability* that a value lies in one distribution rather than the other. For instance, if a boy makes a certain score on a test, and the bi-serial r of that test is known between two curricula, then it is possible from the tables to estimate the *odds* or the *probability* that the boy will be found in one curriculum rather than the other.

TABLE II.—ODDS THAT AN INDIVIDUAL MAKING A SCORE ON A TEST A CERTAIN DISTANCE ABOVE THE MEAN OF ONE GROUP WILL FALL IN THAT GROUP RATHER THAN IN ANOTHER GROUP GIVEN THE BI-SERIAL r BETWEEN THE GROUPS

Bi-serial r	σ above the mean of the higher group								
	−1.00	−.50	.00	+.50	+1.00	+1.80	+2.00	+2.50	+3.00
.05	.....	.....	1.003	1.044	1.087	1.131	1.177	1.225	1.275
.10	.....	.....	1.013	1.097	1.189	1.288	1.395	1.511	1.637
.15	.....	.....	1.029	1.161	1.310	1.478	1.667	1.881	2.122
.20	.....	.....	1.054	1.239	1.456	1.711	2.011	2.364	2.779
.25	.....	.....	1.086	1.332	1.632	2.001	2.452	3.006	3.685
.30	.....	.....	1.129	1.445	1.849	2.366	3.027	3.873	4.956
.35	.....	.....	1.184	1.584	2.119	2.834	3.790	5.070	6.780
.40	.....	.....	1.255	1.757	2.460	3.446	4.825	6.757	9.462
.45	.....	.....	1.344	1.975	2.902	4.263	6.264	9.202	13.52
.50	.....	.....	1.460	2.256	3.486	5.386	8.321	12.86	19.87
.55	.....	.....	1.611	2.626	4.279	6.973	11.36	18.52	30.19
.60	.....	1.051	1.812	3.127	5.393	9.304	16.05	27.69	47.76
.65	.....	1.138	2.087	3.828	7.022	12.88	23.62	43.33	79.47
.70	.....	1.263	2.476	4.855	9.520	18.67	36.60	71.77	140.9
.75	.....	1.446	3.052	6.441	13.59	28.69	60.54	127.8	269.7
.80	.....	1.726	3.956	9.065	20.77	47.60	109.1	250.2	572.6
.85	.....	2.183	5.493	13.82	34.78	87.53	220.3	555.0	1,395
.90	1.068	2.997	8.410	23.60	66.23	185.9	521.6	1,463	4,106
.95	1.457	4.659	14.89	47.61	152.2	486.6	1,553	4,969	15,883

In computing these tables the assumption was made that the number in the two groups is equal, so that p and q each equal .50 in the bi-serial r formula. The method of deriving the tables was to compute the ordinates for both distributions for a given bi-serial r. In deriving the table of odds the ratio between superposed ordinates was computed. For the table of probabilities the ratio of the larger ordinate to the sum of the two superposed ordinates was found.

To be more specific, assume that the two distributions shown in Fig. 3 represent a bi-serial r of .50.

TABLE III.—PROBABILITY THAT AN INDIVIDUAL MAKING A SCORE ON A TEST A CERTAIN DISTANCE ABOVE THE MEAN OF ONE GROUP WILL FALL IN THAT GROUP RATHER THAN IN ANOTHER GROUP GIVEN THE BI-SERIAL r BETWEEN THE GROUPS

Bi-serial r	σ above the mean of the higher group								
	−1.00	−.50	.00	.50	+1.00	1.50	+2.00	+2.50	+3.00
.05	.....	.....	.5008	.5108	.5208	.5308	.5407	.5506	.5605
.10	.....	.....	.5032	.5232	.5431	.5629	.5824	.6018	.6208
.15	.....	.....	.5073	.5373	.5671	.5965	.6251	.6529	.6797
.20	.....	.....	.5131	.5533	.5928	.6312	.6679	.7027	.7354
.25	.....	.....	.5207	.5711	.6201	.6668	.7103	.7504	.7865
.30	.....	.....	.5304	.5910	.6490	.7029	.7517	.7948	.8321
.35	.....	.....	.5422	.6130	.6794	.7392	.7912	.8352	.8715
.40	.....	.....	.5565	.6373	.7110	.7751	.8283	.8711	.9044
.45	.....	.....	.5735	.6639	.7437	.8100	.8623	.9020	.9311
.50	.....	.....	.5935	.6929	.7771	.8434	.8927	.9278	.9521
.55	.....	.....	.6170	.7242	.8106	.8746	.9191	.9488	.9679
.60	.....	.5123	.6444	.7577	.8436	.9030	.9414	.9651	.9795
.65	.....	.5323	.6761	.7929	.8753	.9279	.9594	.9774	.9876
.70	.....	.5581	.7123	.8292	.9049	.9492	.9734	.9863	.99295
.75	.....	.5912	.7532	.8656	.9315	.9663	.9838	.99223	.99631
.80	.....	.6332	.7982	.9006	.9541	.9794	.99091	.99602	.99826
.85	.....	.6858	.8460	.9325	.9721	.9887	.99548	.99820	.999283
.90	.5164	.7498	.8937	.9594	.9851	.99465	.99809	.999317	.999757
.95	.5931	.8233	.9371	.9794	.99347	.99795	.999357	.999799	.999937

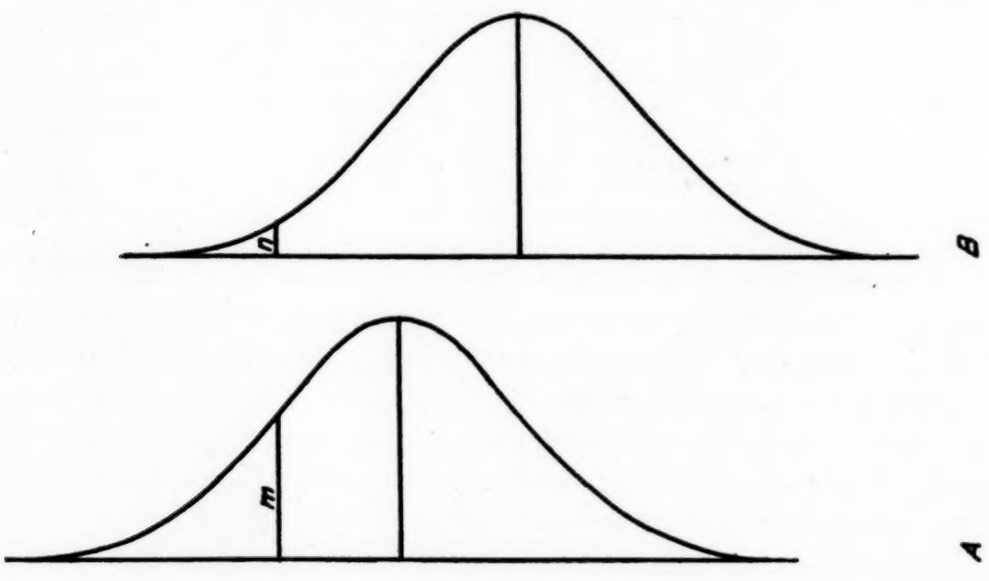


FIG. 3.

If ordinate m is x SD above the mean of distribution A, then it may be found by computation that ordinate n is

$$\frac{2zr}{pq\sqrt{4 - \frac{z^2}{p^2q^2}r^2}} + x$$

SD above the mean of distribution B.[1]

From Table II of Pearson's Tables for Biometricians and Statisticians the height of the two ordinates may be found and the odds and probabilities directly found.

It is the writer's belief that the advantages in using bi-serial r as a measure of overlapping or as a measure of the difference between the means of two distributions has not been fully appreciated. Bi-serial r is easy to find. By comparing the size of bi-serial r with the standard error of bi-serial r, the chances that the difference in the

[1] The determination of the distance in SD units that ordinate n is above the mean of distribution B necessitates the derivation of the following formula. The formula for bi-serial r is

$$r = \frac{M_2 - M_1}{\sigma} \cdot \frac{pq}{z}$$

If

$$M_2 - M_1 = D$$

and

$$\frac{z}{pq} = c$$

we may state that

$$\frac{D}{\sigma} = cr$$

If it be further assumed that $p = q$ and let $d = M_1 - M = M_2 - M$ then $D = 2d$

$$\sigma^2 = \sigma_1^2 + d^2$$
$$= \sigma_1^2 + \frac{D^2}{4}$$

so that

$$\frac{D}{\sqrt{\sigma_1^2 + \frac{D^2}{4}}} = cr$$

by algebraic manipulation it may be found that

$$\frac{D}{\sigma_1} = \frac{2cr}{\sqrt{4 - c^2r^2}} = \frac{2zr}{pq\sqrt{4 - \frac{z^2}{p^2q^2}r^2}}$$

By this formula the difference between the means of the two distributions may be found in terms of the SD of one of the distributions. The distance ordinate n is above the mean of distribution B may be found by adding D to the SD distance of m above the mean of distribution A.

direction indicated is a statistically real difference can be estimated. Bi-serial *r* is comparable to the product moment coefficient of correlation and one can estimate the significance of differences in terms of well-known correlations between tests and other criteria of ability. *It is possible to compare differences between groups when different tests are used because bi-serial r represents a common unit of measurement.* Bi-serial *r* may be used as an effective measure of the results of experimental studies of learning.

STUDIES IN HANDEDNESS: I. A TECHNIQUE FOR TESTING UNIMANUAL HANDEDNESS

R. H. OJEMANN

State University of Iowa

This series of studies in handedness will consist of three articles. In this article, the first of the series, a method of testing unimanual handedness is described. The second article of the series will describe the results obtained in testing bimanual handedness and in studying the relation of bimanual handedness to unimanual handedness. In the third article there will be set forth the results obtained in the investigation of the effect produced upon the speech function by training left-handed individuals to write with the right hand.

The general plan of developing a technique of testing unimanual handedness consisted in obtaining the scores on a series of five unimanual tests for a large unselected group of subjects and in studying the correlation of the scores with a criterion of known handedness.

As a result of a careful study of the literature five tests were selected which appeared promising as unimanual handedness tests. A brief description of the five tests follows:[1]

1. *Ball-throwing Test.*—The subject was given three trials, throwing a ball into a six-inch opening in a target four yards away. A record was made of the hand used in each trial. If the subject used one hand consistently, he was given a final score of *R* or *L* as the case might be. If he used first one hand and then the other he was given a final score of *A*.

2. *Needle-threading Test.*—In this test *S* was asked to thread a needle, using first one hand and then the other. A record was made of the time required for each hand. The difference $L - R$ in seconds was used as the raw score.

3. *Tapping Test.*—An ordinary tapping plate with a metal stylus and an electric counter was used. *S* was given three trials with each hand. The average number of taps per trial was computed and the difference $R - L$ was taken as the raw score.

4. *Paper-cutting Test.*—The task consisted in cutting out, with scissors adaptable to either hand, a design printed on paper. The time required to perform the task with each hand was recorded. The difference $L - R$ in seconds was used as the raw score.

5. *Block-packing Test.*—The subject was instructed to pack forty-nine one-inch wooden cubes into one layer in a box seven and three-sixteenths inches square

[1] Detailed directions for giving each test were worked out and followed. A complete account of the entire study is given in an unpublished Doctor's thesis by the writer on file in the library of the Department of Education of the University of Chicago.

(inner dimensions) and three inches deep. He was directed to use one hand at a time, to pick up not more than two blocks at once, to steady the box with the other hand, and to work as rapidly as possible. The time for each hand was recorded and the difference $L - R$ in seconds was used as the raw score.

The tests were administered to an unselected[1] group of five hundred eighteen pupils in Grades III–VIII inclusive of the public elementary schools of a town of 20,000 inhabitants in a north central state. The distribution of the five hundred eighteen pupils by grade varied from seventy-nine in Grade III to ninety-six in Grade V. There were two hundred fifty-six males and two hundred sixty-two females in the group.

The tests were administered by the writer with the aid of a well-qualified assistant. In each case the testing was done in a separate room where the subject was not disturbed by school activities. One pupil at a time was summoned and tested.

The judgments of parents concerning the handedness of children were selected as the best criterion of known handedness. Parents in general have made more careful observation of their children and have had the opportunity to continue the observation over a longer period of time than have teachers.

It was impossible to secure the parents' judgments for all the five hundred eighteen subjects. A preliminary study of the results on the unimanual tests for the entire unselected group showed that in slightly more than ten per cent of the cases the left hand equaled or excelled the right hand in two or more of the unimanual tests. It was assumed at this point in the study that if these cases were represented on a handedness scale, they would include all the cases falling in the left-hand region, and a sufficient number of cases in the ambidextrous region, so that a point on the handedness scale marking the transition from the left-hand region to the ambidextrous region could be chosen.[2] An attempt was made, therefore, to secure the parents' judgments in all the fifty-two cases in which the raw score was zero or negative on two or more of the five unimanual tests. In all but six of the fifty-two cases complete reports of the parents' judgments were obtained. The incomplete or unsatisfactory reports were due to

[1] The term "unselected" as used in this discussion means a random sample of all the pupils (excluding the pupils in the Negro section) in Grades III–VIII inclusive in the city in which the investigation was made.

[2] Data to be presented later in this report show that this assumption was not an error.

the fact that the parent could not be reached or that the child was under the care of a guardian who had not made sufficiently careful observations.

All the interviews with the parents were made by the writer. In making the interviews each parent was asked to designate whether he considered the child right-handed, left-handed, or ambidextrous. When a satisfactory answer to this question had been obtained, the parent was asked to indicate the basis for his judgment by enumerating some specific activities in which he had observed the child's handedness. An effort was made to have each parent enumerate at least three specific activities.

In order to make clear the nature of the results obtained by interviewing the parents, a detailed discussion of a typical case in which the parent judged the child as left-handed, a typical case in which the parent judged the child as ambidextrous, and of a typical case in which the parent judged the child as right-handed will be given.

Subject 271 was selected as a typical case in which the parent judged the subject to be left-handed. The mother stated that she considered the subject to be left-handed, and that she based her judgment upon the fact that the child ate with the left hand, played ball left-handed, played tennis left-handed, and used a hammer left-handed.

Subject 410 was selected as a typical case in which the parent judged the subject ambidextrous. The mother stated that she considered the child partially left-handed, and that she based her judgment upon the fact that the boy used his left hand to throw a ball and to use a hammer, but that he ate with the right hand, used a pair of scissors with the right hand, and used a knife with either hand.

Subject 110 was selected as a typical case in which the parent judged the subject to be right-handed. The father stated that he considered the child right-handed, and that he based his judgment upon the fact that the child used his right hand in throwing a ball, in eating, in using a hammer, and in combing his hair.

In the above discussion the plans of the investigation and the methods of collecting the data have been described. In the discussion which follows, the results of a study of the data to determine which test or combination of tests correlates highest with the parents' judgments will be given.

The study of the test scores and parents' judgments revealed, in the first place, that none of the five unimanual tests taken separately

is a satisfactory measure of handedness. The tapping test will be selected to illustrate this point. Results similar to those for the tapping test were obtained for the other four tests. The tapping test was selected to serve as an illustration at this point because the data relative to it are typical. It is, therefore, unnecessary to burden the reader with additional tables.

TABLE I.—RELATION BETWEEN THE SCORES ON THE TAPPING TEST AND PARENTS' JUDGMENTS

Score	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−10.5	1	1	1			
− 9.5	0	0				
− 8.5	0	0				
− 7.5	0	0				
− 6.5	3	3	3			
− 5.5	2	2	2			
− 4.5	4	4	4			
− 3.5	7	7	5	1	..	1
− 2.5	8	8	5	2	..	1
− 1.5	6	6	5	1		
− .5	5	3	..	2	1	
.5	25	1	..	..	..	1
1.5	46	4	..	..	3	1
2.5	47	2	..	..	2	
3.5	51	5	1	2	2	
4.5	69	0				
5.5	63	2	..	..	2	
6.5	55	1	..	..	1	
7.5	45	1	..	..	..	1
8.5	39	1	..	..	..	1
9.5	11	1	..	..	1	
10.5	7	0				
11.5	7	0				
.....	..					
.....	..					
.....	..					

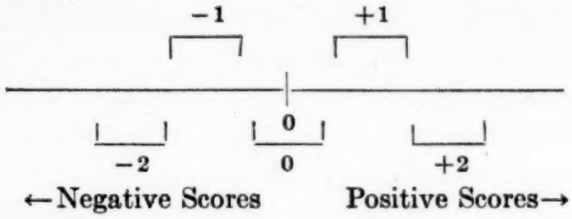
The data relative to the tapping test are presented in Table I. The scores in the first column represent a scale of scores made on the tapping test. The figures in the second column indicate the

frequency with which given scores occurred in the unselected group. The scores in the third column indicate how many cases in a given class interval were investigated. The scores in the first three columns headed "Parents' judgments" indicate the judgments of the parents for the cases investigated. In the column headed "Incomplete data" those cases are recorded for which the parents' judgments could not be obtained, although a serious attempt was made to secure them. The table is to be read thus: In the class interval −6.5 there were three cases. These three cases were investigated and the respective parents reported that they were definitely left-handed cases.

A careful study of the table will show that the tapping test taken alone is not sufficient to separate the left-handed, ambidextrous, and right-handed cases. Ambidextrous cases occur as low as −3.5 and as high as +3.5. In this range a considerable number of left-handed cases also occur. The location of the one case in the +3.5 interval which was judged by the parent to be left-handed is probably not the result of an error in the data. In another phase of this investigation which is to be reported in the third article in this series, a group approximately twice as large as the group used in this part of the investigation was studied. In this larger group four cases were found whose scores on the tapping test were +1.3, +2.0, +3.0, and +5.0 respectively, and who were judged by their respective parents as being left-handed. The tapping test taken alone is, therefore, not sufficient to separate left-handed, ambidextrous, and right-handed cases. Similar data were obtained for the other four unimanual tests.

Since the various tests taken separately did not effect the discrimination desired, it was necessary to make a study of the various combinations of the five tests. This required that the scores on the various tests be reduced to a comparable basis. The method of doing this will now be described. Since on the average nine-tenths of the scores of the unselected group on each of the tests lie on the positive side of zero, the most stable measure of dispersion is the unit of dispersion of the right-handed group. The sigmas of the positive scores of the various tests were found to be as follows: cutting test 64.02, tapping test 3.06, block-packing test 7.13, and needle-threading test 7.02. Instead of computing standard scores, it was judged to be a sufficiently accurate procedure to divide the range of the scores into sigma units and to assign an appropriate value to each score falling within the limits of a given unit. The sigma values given above appeared to be rather large for this purpose and an approximation to a two-thirds

sigma unit was therefore used. For the cutting test a unit of forty was used, for the tapping test a unit of two was used, for the block-packing and needle-threading tests a unit of five was used. The scores on each test were then grouped and values were assigned as indicated in the following diagram:



Scores falling within the range of $-\frac{1}{2}$ unit to $+\frac{1}{2}$ unit were given a value of zero, scores falling within the interval $+\frac{1}{2}$ to $+1\frac{1}{2}$ units were given a value of $+1$, etc. The data showing the values assigned to the raw scores on the different tests are given in Table II.

TABLE II.—RELATION BETWEEN RAW SCORES AND UNIT SCORES

Cutting test		Tapping test		Block-packing and needle-threading tests	
Raw scores	Unit scores	Raw scores	Unit scores	Raw scores	Unit scores
−260.5	−6	−17.0	−8	−32.5	−6
−220.5	−5	−15.0	−7	−27.5	−5
−180.5	−4	−13.0	−6	−22.5	−4
−140.5	−3	−11.0	−5	−17.5	−3
−100.5	−2	− 9.0	−4	−12.5	−2
− 60.5	−1	− 7.0	−3	− 7.5	−1
− 20.5	0	− 5.0	−2	− 2.5	0
19.5	1	− 3.0	−1	2.5	1
59.5	2	− 1.0	0	7.5	2
99.5	3	1.0	1	12.5	3
139.5	4	3.0	2	17.5	4
179.5	5	5.0	3	22.5	5
219.5	6	7.0	4	27.5	6
259.5	7	9.0	5	32.5	7
.....	...	11.0	6		
.....	...	13.0	7		
.....	...	15.0	8		
.....	...	17.0	9		

The method of reducing the scores to a comparable basis could not be applied directly to the throwing test. In order to make it possible to use the scores on the throwing test, the following procedure was

adopted: A preliminary inspection of the scores on the throwing test revealed that the number of individuals who scored *L*, the number who scored *A*, and the number who scored *R* were about the same respectively as the number who made negative scores, the number who made zero scores, and the number who made positive scores on the tapping test. The average unit score of the right-handed group on the tapping test was slightly over +3. This value was then taken as a basis for translating the scores on the throwing test. A score of *R* was given a value of +3, a score of *A* was given a value of 0, and a score of *L* was given a value of −3.

The values resulting from the translation of raw scores into comparable scores will be designated as unit scores throughout this discussion.

The total unit scores for all the various possible combinations of the five tests taken three at a time, four at a time, and five at a time were worked out and the relationships between the total unit scores and parents' judgments were studied. The results of this study will now be presented.

There are ten possible combinations of five tests taken three at a time. In Tables III and IV the data relative to two of these combinations are presented. These two tables were picked from the group of ten as being typical for the group as a whole. In Table III the scores in the first column are the combined unit scores for the three tests, cutting, throwing, and tapping. The scores in the second column indicate the number in the unselected group making a given score. The table is extended far enough to include all the cases that were investigated. All the cases up to and including those making a unit score on the three tests of +3 were investigated. There were five individuals in the unselected group who made a total unit score of +4. Only one of these was investigated. The total in the unselected group making a unit score of +5 was not calculated, but it was more than 5. Five individuals in this group were investigated. The scores in the third column indicate the number in a given interval that were investigated. The scores in the column headed "Parents' judgments" indicate how the parents rated the subjects. The data in the table show that when using the cutting, tapping, and throwing tests together, there is still considerable overlapping between the ambidextrous and left-handed groups.

The data presented in Table IV show that a satisfactory separation of the left-handed from the ambidextrous group cannot be made by

the use of the three tests, tapping, throwing, and needle-threading. The data in Tables III and IV are typical for all the possible combinations for five tests taken three at a time.

TABLE III.—RELATION BETWEEN THE COMBINED SCORES ON THE CUTTING, THROWING, AND TAPPING TESTS AND PARENTS' JUDGMENTS

Total score	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−8	1	1	1			
−7	6	6	6			
−6	7	7	5	..	..	2
−5	5	5	5			
−4	3	3	3			
−3	3	3	2	1		
−2	2	2	2			
−1	2	2	1	1		
0	1	1	..	1		
1	2	2	1	1		
2	2	2	..	1	..	1
3	2	2	..	2		
4	5	1	..	..	1	
5	5+	5	..	1	2	2
6	2+	2	..	..	2	
7	2+	2	..	..	2	
8	1+	1	..	..	1	
9	3+	3	..	..	3	
10	1+	1	..	..	..	1
11	0+	0				
12	0+	0				
13	0+	0				
14	1+	1	..	..	1	
....	...					
....	...					
....	...					

In Tables V and VI data are presented to show the relationship between the parents' judgments and the combined unit scores on two of the five possible combinations on five tests taken four at a time. The data in Table V show that when the scores on the cutting, tapping, throwing, and block-packing tests are combined, the overlapping of the left-handed with the ambidextrous group extends over three intervals. The data in Table VI show that when the scores on the

cutting, throwing, block-packing, and needle-threading tests are combined the overlapping between the ambidextrous and left-handed groups extends over four intervals. These data are typical of the five possible combinations of the five tests taken four at a time.

TABLE IV.—RELATION BETWEEN THE COMBINED SCORES ON THE TAPPING, THROWING, AND NEEDLE-THREADING TESTS AND PARENTS' JUDGMENTS

Total scores	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−11	2	2	2			
−10	2	2	1	..	..	1
− 9	2	2	2			
− 8	2	2	1	..	..	1
− 7	2	2	2			
− 6	7	7	7			
− 5	3	3	3			
− 4	7	7	4	3		
− 3	1	1	1			
− 2	1	1	..	1		
− 1	3	3	1	2		
0	3	3	1	1	1	
1	1	1	..	..	..	1
2	5	2	..	..	1	1
3	9	4	1	1	2	
4	3+	3	..	..	2	1
5	4+	4	..	..	4	
6	1+	1	..	..	..	1
7	2+	2	..	..	2	
....	...					
....	...					
....	...					
....	...					

In Table VII data are given showing the relationship between the total unit scores and parents' judgments when all five tests are combined. This combination effects a more satisfactory discrimination between the left-handed, ambidextrous, and right-handed cases than any of the combinations of three tests or of four tests. The overlapping between the left-handed and ambidextrous groups extends over but one interval. The range of ambiguity between the ambidextrous and right-handed groups extends over but three intervals.

TABLE V.—RELATION BETWEEN THE COMBINED SCORES ON THE CUTTING, TAPPING, THROWING, AND BLOCK-PACKING TESTS AND PARENTS' JUDGMENTS

Total score	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−12	1	1	1			
−11	0	0				
−10	0	0				
− 9	5	5	5			
− 8	2	2	2			
− 7	2	2	2			
− 6	10	10	9	..	..	1
− 5	1	1	1			
− 4	4	4	3	..	..	1
− 3	0	0				
− 2	3	3	2	1		
− 1	2	2	..	2		
0	2	2	1			
1	2	2	..			
2	2	1	..	..	1	
3	8	4	..	1	2	1
4	2+	2	..	..	1	1
5	3+	3	..	1	2	
6	2+	2	..	..	2	
7	2+	2	..	..	1	1
8	2+	2	..	..	2	
9	1+	1	..	..	1	
10	1+	1	..	..	..	1
....	....					
....	....					
....	....					
....	....					
....	....					

The data for the group making a score of +4 and for the group making a score of +5 are not complete, and these facts must be taken into account in interpreting the data of Table VII. However, the data are sufficiently complete to warrant the conclusion that by combining the scores on the five handedness tests the overlapping between the left-handed and ambidextrous groups and between the ambidextrous and right-handed groups is very small and that a satisfactory means of measuring handedness is obtained. The combined score on the

five tests will, therefore, be used in this study to distinguish between left-handed and ambidextrous groups.

TABLE VI.—RELATION BETWEEN THE COMBINED SCORES ON THE CUTTING, THROWING, BLOCK-PACKING, AND NEEDLE-THREADING TESTS AND PARENTS' JUDGMENTS

Total score	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−12	2	2	2			
−11	2	2	1	..	..	1
−10	1	1	1			
− 9	2	2	2			
− 8	2	2	2			
− 7	8	8	7	..	..	1
− 6	6	6	6			
− 5	1	1	..	1		
− 4	1	1	1			
− 3	3	3	3			
− 2	3	3	1	2		
− 1	0	0				
0	3	3	..	..	3	
1	8	5	..	2	2	1
2	5	5	..	2	2	1
3	11	4	..	1	1	2
4	21	2	..	..	2	
5	36	1	..	..	1	
6	1+	1	..	..	1	
....	....					
....	....					
....	....					

In order to make use of the combined scores to investigate problems relating to left-handedness, it is necessary to select the point on the scale which marks the transition from the left-handed to the ambidextrous group. The data presented in Table VII show that all individuals making a score of −4 or lower were judged to be left-handed. The number in the group making a score of −3 is small and it makes little difference, therefore, if −3 or −4 is selected as the upper limit of the left-handed group. Since all subjects making a score of −4 or lower were judged to be definitely left-handed, this score will be chosen as the upper limit of the left-handed group. There were two

TABLE VII.—RELATION BETWEEN THE COMBINED SCORES ON THE CUTTING, TAPPING, THROWING, BLOCK-PACKING, AND NEEDLE-THREADING TESTS AND PARENTS' JUDGMENTS

Total score	Frequency in unselected group	Number investigated	Parents' judgments			
			L	*A*	*R*	Incomplete data
−14	1	1	1			
−13	1	1	1			
−12	2	2	1	..	..	1
−11	3	3	3			
−10	3	3	3			
− 9	4	4	4			
− 8	3	3	2	..	..	1
− 7	5	5	5			
− 6	0	0				
− 5	3	3	3			
− 4	2	2	2			
− 3	2	2	1	1		
− 2	2	2	..	2		
− 1	0	0				
0	1	1	..	1		
1	2	2	..	1	1	
2	5	4	..	2	1	1
3	6	6	..	1	4	1
4	5	1	..	..	1	
5	9	1	..	..	1	
6	1+	1	..	..	1	
7	2+	2	..	..	1	1
8	2+	2	..	..	1	1
9	1+	1	..	..	1	
....	...					
....	...					
....	...					

subjects making a score below −4 for whom the data concerning known handedness were incomplete. Since these two cases appear so far below the upper limit of the left-handed group, it is probably safe to assume that these subjects were left-handed. There were, then, twenty-seven left-handed cases in this group of five hundred eighteen subjects. The proportion of left-handedness is, therefore, $.0521 \pm .007$.[1] The per cent is 5.21. The probable error of this

[1] For the discussion of a probable error of an observed proportion see Holzinger, K. J.: "Statistical Methods for Students in Education." Pp. 248–250.

proportion is relatively large and this fact must, of course, be taken into account in interpreting it.

The distribution of the total unit scores for the combined five tests for the entire group is given in Table VIII and is presented graphically in Fig. 1. A brief study of the graph shows that the curve is distinctly bimodal. There are relatively few cases in the interval the mid-point of which is zero, and the curve rises as one

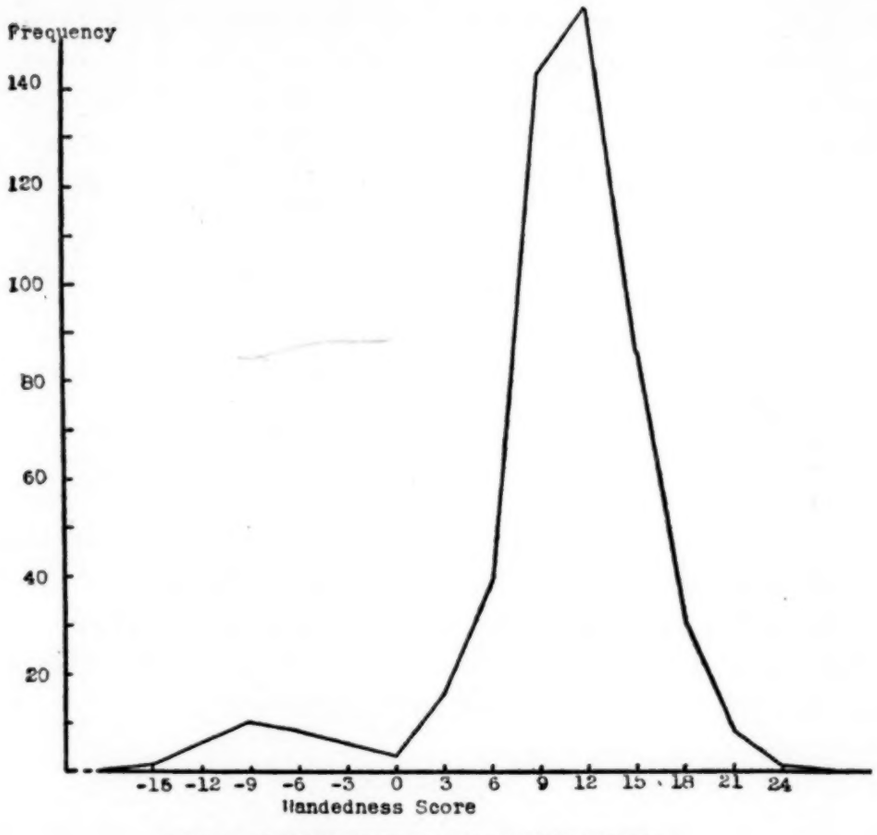


Fig.1 Distribution of Total Unit Scores

moves toward either end of the scale. The rise on the right-hand side is, of course, much greater than that on the left, but the bimodality is clearly evident.

The distribution curve for handedness is, therefore, not of the unimodal normal frequency type. There are relatively few individuals who may be classed as ambidextrous. The tendency is in either the left-hand or right-hand direction. The results show that handedness in one unimanual activity tends to be correlated with handedness in other unimanual activities. It appears, therefore, that some sort of a unifying factor is present.

TABLE VIII.—DISTRIBUTION OF THE TOTAL UNIT SCORES ON THE FIVE UNIMANUAL TESTS IN THE UNSELECTED GROUP

Total unit score	Frequency	Total unit score	Frequency
−14	1	5	9
−13	1	6	10
−12	2	7	20
−11	3	8	38
−10	3	9	43
− 9	4	10	65
− 8	3	11	57
− 7	5	12	57
− 6	0	13	43
− 5	3	14	37
− 4	2	15	25
− 3	2	16	24
− 2	2	17	11
− 1	0	18	10
0	1	19	10
1	2	20	5
2	5	21	1
3	6	22	2
4	5	23	1

It is difficult to explain in the light of this finding how handedness can be entirely socially conditioned, as has been proposed. There is no environmental factor that will account for the distinct rise to a new mode at the left-hand side of the handedness curve.

SUMMARY

In this article a method of testing unimanual handedness is described. Three important conclusions may be drawn from the data.

1. A single test such as the tapping test, the ball-throwing test, the paper-cutting test, the needle-threading test, or the block-packing test cannot be used to differentiate accurately between the various unimanual handedness groups. On each test there is considerable overlapping of the scores made by the left-handed individuals with the scores made by the ambidextrous subjects, and similar overlapping of the scores made by the ambidextrous subjects with the scores made by right-handed individuals.

2. If, however, the scores on an increasing number of tests are combined, the overlapping of the scores tends to decrease. When the

scores on the five tests—tapping, paper-cutting, ball-throwing, block-packing, and needle-threading—are combined, the overlapping of the scores made by the left-handed, ambidextrous, and right-handed individuals is relatively small.

3. When the combined scores on the five unimanual tests for an unselected group are presented graphically, the resulting curve is distinctly bimodal. Handedness appears, therefore, to be a general factor which tends to divide individuals into two groups. Relatively few individuals are ambidextrous. Individuals tend to cluster about a right-handed tendency or a left-handed tendency. This finding tends to indicate that handedness is an inherited characteristic. There is no environmental factor which will account for the bimodal distribution of handedness found in this study.

INTELLIGENCE OF COLONY PEOPLE

W. W. LUDEMAN AND J. R. McANELLY

Education Department Southern State Normal School, Springfield, S. D.

In the various mental tests given by the authors to pupils in city and rural communities with varying conditions of environment, those pupils with the richer environmental experience have invariably shown a higher degree of intelligence. This has been true where pupils were of the same physical status and lineage, and naturally has given rise to the question as to how much of the IQ, as found by the average mental test, is innate capacity and how much acquired intelligence.

There has been considerable interest for some time as to the effect of a very limited environment upon the ability of a person to be successful in standardized intelligence tests. Two questions are dominant in the minds of many students of the measurement of intelligence: (1) Is there any difference between innate capacity and acquired intelligence? and (2), Do mental tests really measure the innate or the acquired intelligence?

As a step toward answering these questions, the authors had an opportunity to study a group of thirty-two Communistic colony children whose environment is extremely limited.

This colony is based upon religious traditions, and is so organized that everything is done in common. Its members are averse to modern progress, and to maintain that principle they deny themselves all modern conveniences and keep their children aloof from contact with the outer world. Their habits and customs are those of past generations. They have no telephones, no electric lights, no newspapers. There is not an automobile in the colony. As an agricultural community they are thrifty and prosperous, thus showing that they are possessed of native intelligence. The business affairs of the colony are managed by an overseer who handles all income and expenditure. In this way no opportunity is given the members, and especially the children, for handling money or for conducting business affairs of any kind. Children are not allowed to leave the colony, and adults only once a year. Their church, schoolhouse, and meeting-house are the same one-room building. Marriages are arranged for the most part by the parents.

The school from which this group of thirty-two pupils was taken is composed of eight grades and is conducted in the one-room building

with two men teachers in charge. The head teacher is well qualified, holding a South Dakota state certificate, while his assistant, a member of the colony, holds no certificate at all. The children look well cared for and appear to be as happy as any average group, even though play is not encouraged and there is absolutely no playground equipment around the school building.

Three different mental tests were used in making this study. The Kingsbury Primary Group Intelligence Scale, Form A, and the Meyers Mental Measure were used in Grades II to IV. These tests consist wholly of pictures, and are independent of reading and writing ability. For Grades V to VIII the Meyers Mental Measure and the National Intelligence Test, Scale A, Form I, were used. The National Intelligence Test is adapted to these grades and is of the army type, while the Meyers Mental Measure is adapted to all grades, hence it was used for both groups.

TABLE I.—IQ'S FOR GRADES II–IV

Pupil number	IQ on Kingsbury	IQ on Meyers
1	85	85
2	71	73
3	71	71
4	72	70
5	83	67
6	51	63
7	76	74
8	90	80
9	78	76
10	55	58
11	71	75
12	70	75
13	71	71
Median.....	74.5	74
Mean.....	72.6	72.1

TABLE II.—IQ'S FOR GRADES V–VIII

Pupil number	IQ on National	IQ on Meyers
1	70	70
2	66	64
3	77	70
4	60	60
5	75	60
6	78	60
7	62	58
8	70	80
9	70	90
10	60	70
11	60	80
12	69	69
13	65	70
14	70	71
15	66	65
16	65	73
17	40	58
18	97	92
19	40	48
Median.....	67	67.5
Mean.....	66.3	68.8

In Table I are shown the comparatively low intelligence quotients of the younger group of children. The median intelligence quotient

on the two tests closely agree, the median on the Kingsbury test being 74.5 while that of the Meyers test is 74. The mean intelligence quotient on the Kingsbury test is 72.6 while on the Meyers test it is 72.1, thus showing a marked consistency in reaction to the two tests and seeming to indicate that the results secured on these pupils are fairly accurate and show the true IQ according to the tests.

In Table II is shown a slightly lower median intelligence quotient for the older group than that shown by the younger group in Table I. However, there is the same consistency of agreement in reaction between the National test and the Meyers test with this older group that is shown between the Kingsbury and the Meyers test with the younger group. This agreement in reaction throughout would indicate that the tests were given consistently and with fairness to both groups, and that the results obtained are quite exact.

There were a number of marked instances during the tests where pupils gave evidence of a lack of environmental interpretation, such as writing a word in a blank to complete a sentence. Following are a few illustrations from the tests given: "Mother is ———— doughnuts," "Twenty-five cents is ———— of a dollar." In the former a number of the pupils filled in the blank by using the word "picking." In the latter instance the blank was filled by such words as "nickel," "dot," "doda," showing an utter lack of acquaintance with the data asked for. In an instance where it was required to place a cross below an object the children did not know what a cross was. In another part of one test they were required to select a church from a group of buildings in a picture. The distinguishing feature of the church was its steeple, but instead of underlining this they underlined the picture of the box-shaped factory building that looked something like their own church and meeting-house.

The part of the test in which these colony pupils showed the greatest aptitude was on matching symbols where no environmental interpretation was necessary. This fact is plainly emphasized in test two and test five of the National Intelligence Scale. In test two of the National scale, where the questions are completion exercises requiring an acquiring type of intelligence, the pupils answered correctly one hundred fifteen out of a possible three hundred eighty, or thirty per cent. In test five, where the pupil is required to distinguish symbols, thus showing the innate capacity of an acute observation, they performed correctly one thousand seventeen out of a possible two thousand two hundred eighty, or forty-nine per cent.

The opportunity to test these colony children firmly convinced the authors of this paper that the limited environment of the pupils was a decided handicap in taking the tests. Many of the questions, name-words, and problems had never been real experiences with these children, and hence they had had no chance to form definite reactions toward them. Such handicaps rather than any lack of innate capacity are responsible for the low average intelligence quotients, as is emphasized by the fact that in those parts of the tests where references to worldly environmental conditions were not brought in, the colony children performed as well as normal children do.

Conclusion

According to the above data the following conclusions are suggested:

1. The median of intelligence of both groups of colony children as measured by the intelligence tests is much lower than the native capacity of the average group of boys and girls.
2. Both groups experienced the greatest difficulty in the tests where environmental knowledge was required, and least difficulty where native intelligence was required.
3. The low median in both groups is due to limited environmental conditions rather than to a low native capacity.
4. The average mental test is a measure of acquired intelligence as well as innate capacity.

WHEN DOES A TEST MEASURE THE SAME FUNCTIONS AT ALL LEVELS?

BARBARA S. BURKS

Stanford University

Considerable fluctuation sometimes occurs in children's scores (relative to the norms for their ages) when a test is repeated, or when a second form of a test is administered, after a lapse of several years. It is a matter of great practical and theoretical significance whether such apparent changes in ability correspond to actual changes in the abilities of children, or whether they reflect merely the nature of the test which is used as a measuring instrument.

On such a test as the Stanford-Binet there is almost no overlapping of the actual material used to test a six-year-old child with that used to test, let us say, a twelve-year-old child. When a child first tested at age five or six is later tested at age eleven or twelve, he is being measured on the two occasions by virtually two different tests. Unless there is complete overlapping of the mental functions that are tapped at the two portions of the scale, the child's IQ may appear to have suffered a change even though his real mental caliber has remained constant throughout the interval between tests.

Correlations between tests and retests on the Stanford-Binet tend to decrease a little as the interval between testings increases. This fact has been thought by some to constitute good evidence that the test is measuring at the higher levels functions that differ to a certain extent from those measured at the lower levels. It would be equally pertinent, however, to ask if the trend in the correlations between tests and retests might mean, instead, that the mental growth of children is occasionally subject to spurts, plateaus, or even permanent changes in rate. There is nothing in the correlations themselves that enables us to distinguish between factors inherent in the test and factors inherent in the children tested.

A method is developed in this paper for determining when fluctuations in IQ (or in any test score) over a period of time are due to real changes in ability and not to changes in the nature of the measuring instrument, and for ascertaining when the fluctuations may be accounted for entirely by changes in the functions measured.

The method requires the use of a rather large group of sibling tests. It is clear that if a test measures at one level functions that are to

some extent independent of the functions measured at another level, resemblances between siblings will have a better chance to show up if both members of each sibling pair are tested when at or near the same age. We should of course expect a higher correlation between children's height and the height of their siblings than between children's height and some quite different trait in their siblings, such as strength of grip. Similarly, we should expect a higher correlation between

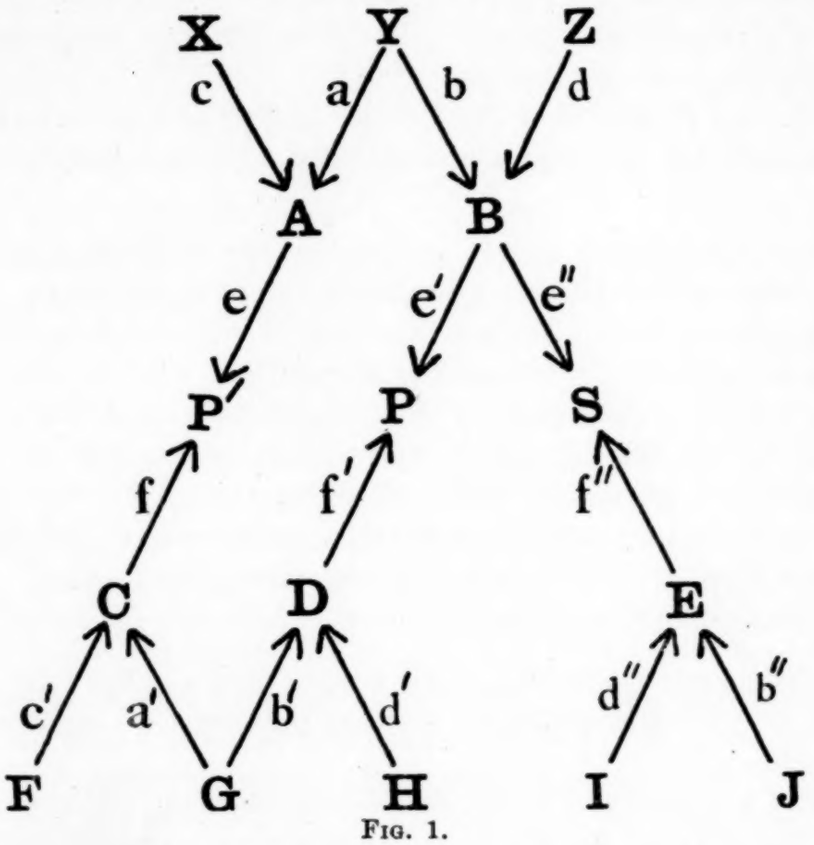


FIG. 1.

mental test scores of siblings when the same, rather than partly disparate, mental functions were being tapped.

If progressive changes in test scores (for the sake of concreteness let us say IQ's) beyond the fluctuations predictible from the reliability coefficient of the test are due entirely to changes in the nature of the functions measured, what effect should the phenomenon have upon sibling correlations? Wright's method of path coefficients[1] provides a means for calculating this effect.

[1] Wright, Sewell: Correlation and Causation. *Journal Agricultural Research*, Vol. XX, 1921, pp. 557–585. The Theory of Path Coefficients. *Genetics*, Vol. VIII, 1923, pp. 238–255.

The situation is shown diagrammatically in Fig. 1. The arrows represent the paths through which contributions are made to the IQ variance of:

P' = propositi at a given age
P = propositi at a different age
S = siblings of propositi at same age as P.

The small letters by which the arrows are designated represent the values of the path coefficients.[1] Paths of the same letter have equal value regardless of superscripts.

X, Y, and Z, and their dependents, A and B, are variable contributions made by heredity or environment to both members of sibling pairs.

X is measured by a test at one age but not at another age.
Z is measured at the second age but not at the first age.
Y is measured at both ages.
A is completely determined by X and Y.
B is completely determined by Y and Z.

F, G, H, and their dependents, C and D, are contributions to the IQ of propositi but not to the IQ of siblings.

F is measured by our test at the first age but not at the second age.
H is measured at the second age but not at the first age.
G is measured at both ages.

I, J, and their dependent, E, are contributions to the IQ of siblings, but not to the IQ of propositi, and are measured by our test at the second age.

Wright has shown that the correlation between any two variables may be found by taking the products of the path coefficients summed over all the paths of influence by which the variables are connected.

Accordingly we may set down the following equations:

$$r_{P'S} = eabe'' \quad (1)$$

$$r_{PP'} = eabe' + fa'b'f' \quad (2)$$

$$r_{PS} = e'e'' \quad (3)$$

$$e^2 + f^2 = 1 \quad (4)$$

[1] A path coefficient is the proportion of the standard deviation of a variable which would remain if all influences except the one in question were kept constant *while the influence in question varied as much as before.* It is by virtue of the clause in italics that a path coefficient differs fundamentally from a partial correlation coefficient.

Since path coefficients designated by the same letters have equal values regardless of superscripts, we may simplify the above equations as follows:

$$r_{P'S} = abe^2 \tag{5}$$
$$r_{PP'} = abe^2 + abf^2 \tag{6}$$
$$r_{PS} = e^2 \tag{7}$$
$$e^2 + f^2 = 1 \tag{4}$$

Now

$$\frac{r_{P'S}}{r_{PS}} = ab \text{ by (5) and (7)}$$

and

$$r_{PP'} = ab \text{ by (6) and (4)}$$

Hence

$$r_{PP_1} = \frac{r_{P'S}}{r_{PS}},\ r_{PS} = \frac{r_{P'S}}{r_{PP'}},\text{ and } r_{P'S} = (r_{PS})(r_{PP_1}) \tag{8}$$

These equalities refer, of course, to correlation coefficients corrected for attenuation, since the random errors of a fallible test would invalidate the theoretical relationships represented in Fig. 1.

Since r_{PS} exceeds $r_{P'S}$ when the deviation from unity of $r_{PP'}$ is accounted for by a change in the functions measured by a test, it follows that if $r_{P'S}$ is equal to r_{PS}, the test is measuring the same functions at the two levels with which we are concerned, and changes in IQ (beyond those due to the unreliability of the test) are attributable to actual changes in children's ability.

It does not follow with certainty, however, that whenever r_{PS} does exceed $r_{P'S}$ the deviation of $r_{PP'}$ from unity is due in whole, or even in part, to changes in the nature of the functions measured. If fluctuations in the IQ's of children should be systematic instead of random, and if, further, there should be a correlation between siblings with respect to the ages at which such fluctuations took place, we might still expect r_{PS} to exceed $r_{P'S}$. To take the limiting case, if the fluctuations in sibling IQ's were perfectly correlated with the fluctuations in IQ's of the propositi (which of course would be almost inconceivable), the disparity between r_{PS} and $r_{P'S}$ would be as large as though the deviation of $r_{PP'}$ from unity were attributable entirely to changes in the nature of the test. If in a series of experimental data r_{PS} should prove to differ from $r_{P'S}$ it would therefore be a wise precaution to investigate the fluctuations in IQ's of siblings and propositi for a possible correlation, even though this would greatly increase the bulk of test data required to study the problem.

Table I gives theoretically predicted values for $r_{P'S}$ corresponding to several assumed values of $r_{PP'}$ and r_{PS} on the hypothesis that changes in the functions measured account for the deviation of $r_{PP'}$ from unity. Assumed correlation coefficients corrected for attenuation are substituted in formula 8. The "raw" correlations are derived from the "corrected" correlations from the formula, $r_{SW} = r_{12}/\sqrt{r_{11}r_{22}}$ upon the assumption that the reliability coefficient of the test is equal to .90.

TABLE I

	Assumed $r_{PP'}$	Assumed r_{PS}	Predicted $r_{P'S}$
Corrected for attenuation	.95	.55	.52
Raw	.86	.50	.47
Corrected for attenuation	.90	.50	.45
Raw	.81	.45	.40
Corrected for attenuation	.85	.50	.42
Raw	.76	.45	.38
Corrected for attenuation	.80	.55	.44
Raw	.72	.50	.40
Corrected for attenuation	.80	.50	.40
Raw	.72	.45	.36
Corrected for attenuation	.75	.55	.41
Raw	.68	.50	.37

A FURTHER NOTE ON THE CORRECTIONS OF CERTAIN ERROR FORMULAS

HARL R. DOUGLASS

University of Minnesota

During the academic year of 1927–1928, the author experienced several occasions to employ a formula for determining the standard error of a coefficient of correlation estimated by means of the Spearman-Brown prophecy formula. He found in the literature two formulas offered for that purpose. One of these was furnished by Shen in the October, 1924, issue of this Journal. The second, offered by Holzinger and Clayton in the May, 1925, issue of the same journal, was claimed by its authors to yield more accurate values due to the neglect of certain higher order terms in the derivation of Shen's formula. While Shen had offered in the February, 1926, number of this Journal rather convincing empirical proof of the superiority of his formula, the present author was not entirely satisfied and attempted to verify the mathematical proof of both formulas. This he was able to do for Shen's formula. As stated in an article appearing in the September, 1929, issue of this Journal, he was able to verify the Holzinger-Clayton formula only by a procedure which was definitely mathematically not permissible. In that article he was able also to demonstrate quite clearly that the Holzinger-Clayton formula could not but yield values which were greater deviations from the true values of the standard error than values obtained by use of the Shen formula.

As stated in so many words, the object of the note published in the September issue was "to investigate the relative accuracy of the two formulas." As a tentative formula thought to be a closer approximation than either of the other formulas until further investigation would show, if possible, that such was not true, the author appended a formula based upon the assumption that the error resulting in r estimated by means of the Spearman-Brown formula resulting from an error equivalent to the standard error of the coefficient upon which the estimate would be based, would be a very close approximation to the true standard error of the estimated coefficient.

This formula was made the subject of a very severe criticism by Holzinger in the December number of this Journal. Holzinger writes: "Douglass gives a new formula which he says 'will furnish values

which, in at least most situations to which the Spearman-Brown formula would be applied' gives a still closer approximation than (3),[1] or a formula derived by the writer. These remarkable statements are not only entirely unsupported, but the method used in the proof is inapplicable, the application made has a serious algebraic error, and finally the result is said to approximate the worst formula of the lot." The algebraic error consists in writing or printing $(1 - r^2)N^{\frac{1}{2}}$ instead of $(1 - r^2)^2N^{-\frac{1}{2}}$. In pointing out this inaccuracy, Holzinger renders a service. The formula is correctly stated in Holzinger's article. To have characterized it as a "serious algebraic error" rather than as a clerical or typographical error was somewhat ungenerous, in view of the fact that to anyone who traces the derivation of the formula, the former possibility must appear quite remote. Yet the fault, from whatever source, was unfortunate and it is well that Holzinger called attention to it.

The present author sees nothing very "remarkable" in the statements he made. That they were "entirely unsupported" is hardly true, as reference to page 437 of the article in the September, 1927, issue will reveal. It is quite possible that the formula is the "worst formula of the lot," if that is what Holzinger intended to say. Yet, it seems almost impossible that it is worse than the Holzinger-Clayton formula, which cannot be correctly derived and which is certain to result in values further from the true values than those yielded by the Shen formula.

Tabulated below are given values obtained by use of Shen's and the present author's formulas for different values of N, n, and r.

$N = 16$			
n	$r = .3$	$r = .6$	$r = .9$
.5	S = .158	.085	.0785
	D = .160	.0857	.0786
.2	S = .260	.125	.03006
	D = .277	.126	.03009
5.0	S = .235	.0691	.011224
	D = .283	.0695	.011229
10.0	S = .166	.039	.00573
	D = .239	.041	.00574

$N = 36$			
n	$r = .3$	$r = .6$	$r = .9$
.5	S = .104	.0566	.0523
	D = .106	.0568	.0524
.2	S = .178	.0833	.017544
	D = .181	.0835	.017547
5.0	S = .156	.0447	.00748
	D = .163	.0468	.00749
10.0	S = .110	.0059	.00460
	D = .132	.0063	.00461

[1] Shen's formula.

$N = 100$

n	$r = .3$	$r = .6$	$r = .9$
.5	S = .0629	.03400	.03140
	D = .0631	.03404	.03141
.2	S = .1076	.050	.012018
	D = .1078	.05008	.012019
5.0	S = .094	.027	.0044895
	D = .096	.028	.0044899
10.0	S = .066	.015	.00276
	D = .079	.017	.00277

These tabulated values bear out the statements made concerning the formula presented by the author at that time in the September issue of the Journal that "this formula, like (5) gives values greater than those yielded by the Shen and the Holzinger-Clayton formula and only when n is large and N and r small will values yielded by it differ greatly from those yielded by the Shen formula."

It will also be noted from the tabulations that the values obtained by use of the author's formula are, as they should be, typically larger than those obtained by use of Shen's formula, and materially greater only when n is large and N and r small, conditions under which the Spearman-Brown formula would rarely be employed.

As to the inapplicability of the method of proof, Holzinger presents a very convincing criticism. However, it should be remembered that the proof offered in the September Journal was based upon an assumption which was known to be only a close approximation. That application of the assumption would stand such rigorous test as proposed by Holzinger, namely, generalizing the derivation and then substituting arbitrary values, is not surprising. The author at that time made no claims with respect to the validity of the formula in a more general form and does not now do so. However, in view of the very close approximation of the values yielded by Shen's formula to the values which the complete formula would certainly yield and of the very small difference between values obtained by the use of Shen's formula and those yielded by the formula under discussion, it would seem that Holzinger's suggestion that Shen's formula should be used as the best yet available for the value of the standard error of a coefficient of correlation estimated by means of the Spearman-Brown formula will lead no one into very serious error.

As previously stated it was the object of the author's previous article to investigate the Shen and the Holzinger-Clayton formula and

to demonstrate the fact that the latter formula would yield not more but less accurate values than the former. Holzinger now states that the Holzinger-Clayton formula was "obviously" wrong. Nevertheless, published in May, 1925, it was allowed to stand in the literature without correction or retraction until its error, which after all was not so obvious, was pointed out by the present author in the September Journal.

NOTE ON THE STANDARD RESPONSE ERROR IN A MEASURE OF IMPROVEMENT

JACK W. DUNLAP AND EDWARD E. CURETON

Territorial Normal and Training School, Honolulu, Hawaii

In the March issue of this Journal, Lindquist gives the standard response error of a measure of improvement obtained by subtracting the initial score on a fallible test from the final score (after instruction or practice) on the same test or one comparable to it. He proposes

$$\bar{\sigma}_g^2 = \bar{\sigma}_a^2 + \bar{\sigma}_b^2 - 2r_{ab}\bar{\sigma}_a\bar{\sigma}_b \tag{1}$$

where a is an initial score

b is a final score
g is a gain in score, so that $g = b - a$
$\bar{\sigma}_a$ is the standard error of an initial score
($\bar{\sigma}_a = \sigma_a\sqrt{1 - r_{aA}}$, r_{aA} being the initial reliability)
$\bar{\sigma}_b$ is the standard error of a final score

This formula is in error. The last term should not be present. A different derivation will make this clear.

Let $a = \alpha + \delta$, and let $b = \beta + \Delta$, where a and b are the initial and final scores as measured, α and β are the corresponding true abilities, and δ and Δ are the errors of measurement.

Then

$$g = \beta + \Delta - \alpha - \delta$$

and

$$\bar{\sigma}_g^2 = \sigma_\beta^2 + \sigma_\Delta^2 + \sigma_\alpha^2 + \sigma_\delta^2 + \text{product terms}$$

If we make the usual assumption that the errors of measurement are uncorrelated with each other and with the true abilities, all product terms vanish with the exception of one, $r_{\alpha\beta}\sigma_\alpha\sigma_\beta$. But by definition α and β are true abilities of the individual and do not vary with repeated applications of the tests. Therefore σ_β^2, σ_α^2 and $r_{\alpha\beta}\sigma_\alpha\sigma_\beta$ are all equal to zero, whence

$$\bar{\sigma}_g^2 = \sigma_\Delta^2 + \sigma_\delta^2 = \bar{\sigma}_b^2 + \bar{\sigma}_a^2$$
$$\bar{\sigma}_g = [\sigma_a^2(1 - r_{aA}) + \sigma_b^2(1 - r_{bB})]^{\frac{1}{2}} \tag{2}$$

If we assume that $\sigma_a = \sigma_b$ and $r_{aA} = r_{bB}$; *i.e.* that the initial and final tests are strictly comparable,

$$\bar{\sigma}_g = 1.414\sigma_a\sqrt{1 - r_{aA}} \tag{3}$$

which is identical with Lindquist's formula (7). All his other approximations suffer from the same error as (1), which is due apparently to some confusion of the individual response error with the sampling error.

CORRECTIONS

In the article by Keys and Whiteside on "The Relation of Nervous-emotional Stability to Educational Achievement" which appeared in the September issue, the statement in conclusion at the bottom of p. 440 that the nervous and emotional children averaged "nearly two *months* lower in mental and educational age" should read "nearly two *years* lower." In the reproduction of the rating scale on p. 432, horizontal lines running directly above the numerals in parentheses were accidentally omitted.

In the article "A Comparison of Difficulty and Accuracy in the Learning of Bright and Dull Children in a Motor-memory Task" by Professor Frank T. Wilson, which appeared in the October issue, an error occurred in the first paragraph. The sentence reading "It appeared from that study that for groups of children nine and twelve years of age, half of each group at a level of seventy to eighty IQ . . . " etc., should have read *eighty to ninety IQ*.

A REPLY TO PROFESSOR SPEARMAN

HAROLD DEAN CARTER

University of Minnesota

Since scientific controversies frequently lead to clarification of issues, a comment on Professor Spearman's recent criticisms[1] of my attempt to analyze the nature of mechanical ability[2] may be in order.

His main criticism was directed against my use of the intercolumnar correlation method as a criterion for the presence of the familiar "two factors," *g* and *s*. He characterizes this method as obsolete, yet in his book, "The Abilities of Man," on p. 139 he has a table of twenty series of data which were analyzed by this method, and the results were used in support of his theory. This suggests that even though the method may now be labeled as "obsolete," nevertheless it is permissible to continue to report results secured by that method. Furthermore, it is to be remembered that the newer tetrad-difference method of analysis mathematically expresses the same fundamental relations. In my own work I recognized this fact and accordingly analyzed my data by means of both methods, discovering that the trend of results was the same. This was inevitable in the very nature of things. Since I employed the tetrad-difference method in addition to the intercolumnar correlation method, I do not believe that Professor Spearman was justified in his main criticism.

Mr. Spearman also criticised my work by stating that my tables of correlations did not satisfy the only condition under which the intercolumnar criterion is applicable. It is true he has a criterion for the exclusion of certain pairs of columns, which is arbitrary and not wholly free from criticism. After reading Thomson's mathematical criticisms[3] of the "exclusion criterion" I did not regard this criterion as essential, but nevertheless I decided to apply it to my main table of intercorrelations. It was my misfortune, however, to take this criterion directly from Professor Spearman's 1914 paper,[4] in which

[1] Spearman, C.: A Truce to "Barking In." *Journal of Educational Psychology*, Vol. XXI, 1930, pp. 110–111.

[2] Carter, Harold D.: The Organization of Mechanical Intelligence. *Journal of Genetic Psychology*, Vol. XXXV, 1928, pp. 270–285.

[3] Thomson, G. H.: On the Degree of Perfection of Hierarchical Order among Correlation Coefficients. *Biometrika*, Vol. XII, 1919, pp. 355–366.

[4] Spearman, C.: The Theory of Two Factors. *Psychological Review*, Vol. XXI, 1914, pp. 101–115. See p. 112.

he himself misstated it. I should have secured the method as correctly stated by Spearman in his 1912 paper.[1] The criterion, therefore, was improperly applied, which led to my incorrect statement that all columns of my Table III satisfied it. But recently, after reading his criticism, I applied the criterion again—this time correctly. Of the eight columns in Table III, four satisfied it and four did not. The average correlation between all columns in the table was .41. The average of the six intercorrelations of the columns which did satisfy the criterion was .39. For all other pairs of columns the average intercorrelation was .42. Correct application of the criterion is thus shown to have no important bearing on the results.

Professor Spearman also criticised my paper because his own references to one published article and to two unpublished reports in his "The Abilities of Man" were neglected. He claims that those investigations reveal mechanical ability to be unrelated to *g* and to involve a group factor. But my analysis showed overlapping group factors. It is to be noted that my results differ from those quoted by Spearman in emphasizing *overlapping group factors* rather than *a single group factor*. Thus it would appear that the Minnesota Mechanical Abilities Research Project makes an additional contribution to previous knowledge dealing with the abilities of man.[2]

---

[1] Spearman, C. and Hart, B.: General Ability, Its Existence and Nature. *British Journal of Psychology*, Vol. V, 1912, pp. 51–84. See p. 56.

[2] Paterson, D. G., Elliott, R. M., Anderson, L. Dewey, Toops, H. A., Heidbreder, E.: "The Minnesota Mechanical Ability Tests." The University of Minnesota Press, 1930, pp. 1–640. See especially Chapter II, The Theory of Unique Traits, Chapter XI, The Organization of Mechanical Ability, and Chapter XII, Mechanical Ability as a Unique Trait.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

The Construction and Validation of an Art Test, by Margaret McAdory. New York: Bureau of Publications, Teachers College, Columbia University, 1929. Pp. 35.

Scientific measurement is being carried more and more into the field of the graphic arts. Since much emphasis has heretofore been placed upon the measurement of recognition of compositional excellence in black and white, Miss McAdory's test is a contribution in that it is the second and the better of two tests, now available, which include specific problems in color.

Selection of items and their order of merit were determined by judgments of one hundred individuals drawn from the five general groups of art patrons—producers, teachers, critics, buyers, and "competent lay critics." Inasmuch as the judges composed a group representing many and various points of view in the arts, the criteria for selection of items are probably as valid as, and perhaps more valid than, those employed in the construction of other tests in this field. A valuable feature of Miss McAdory's test is that it includes a wide variety of problems.

The test is designed for use in the upper elementary grades and above. It is regrettable that no adequate norms are yet available.

Miss McAdory's thesis is confined largely to a discussion of the devising of a reliable method of scoring ranked preferences. After a trial and error examination of twenty-one different techniques, the method of scoring one point for each correct placement was adopted. This yielded a coefficient of reliability of .93. Two rather obvious techniques were overlooked which would have involved far less work for the author. One of these would have been the weighting of deviations from the correct order of items, within a particular series, on the basis of the amount of deviation found in judgments by experts of that series. This method would have required the giving of differential weights to each problem in the test. The second, and probably the more reliable, would have been a procedure similar to the one worked

out by Mr. Howard E. Wilson and reported in the *School Review* of February, 1930. The latter not only gives credit for the correct placement of an item within a series, but also credits the order of items, even though no correct placement may have been made. The method recommended by Miss McAdory takes no account of this order of items.

MARCIA E. MENDENHALL.

Bellevue Hospital New York.

Social Control of the Mentally Deficient, by Stanley Powell Davies. New York: Crowell Pub. Co., 1930. Pp. VIII + 389.

The aim of the book is to present the social aspects of mental deficiency: "To show how the various phases of public opinion and action with reference to the social control of this problem have followed the evolution of scientific thought and knowledge concerning mental deficiency." (P. VII.)

In brief but interesting manner the author traces in the first four chapters the beginnings made in locating and caring for the feebleminded. In like manner he disposes of the alarms sounded by the eugenists, the indictment of the feebleminded as potentially criminal, the flurry caused by the advocation and practice of sterilization, and segregation and its disappointing results. Chapter X, "Changing Concepts of Heredity," shows how ideas of heredity have changed, and agrees in substance with Jennings' contention that the result from the combination of the genes of two parent individuals under the influence of varying environmental conditions cannot be accurately predicted. From data gathered he concludes that "half or less of the mental deficiencies found among the institutional group and possibly an even smaller proportion of the mental deficiencies discovered in the community are of the familial or heredity type." (P. 163.) The fertility of the feebleminded is the fertility of people living on a low cultural and economic plane, and unsophisticated in the trends of modern society. In short, the whole question of the rapid propagation of the feebleminded and the alleged swamping of civilization by this defective stock "belongs to the legend of the feebleminded"; and the "only one thing of which we can be certain at present" concerning the inheritance of feeblemindedness "is of our uncertainty." (P. 167.)

Chapter XI points out that the mentally deficient contribute more than their due proportion of social offenders, not because of the direct relationship between a low IQ and social misbehavior, but because they have come for the most part from homes lacking in good parental

guidance, decent housing, adequate economic resources, and a wholesome neighborhood environment. The author discusses, with ample case histories, the advantages of the institution as a socializing force, the colony as an outlet for man power, and the girls' colony as a proving-ground. Chapter XV maintains that when our public schools fully recognize that they have no right to deprive a child of educational advantages suited to his low level of intelligence, and when they provide him with competent psychological, psychiatric, and visiting-teacher service, a large part of the problem of mental deficiency will have been solved. The work of the school, however, should of necessity be supplemented and coordinated with community supervision.

Chapter XVII further develops the idea that behavior is a process of the interaction of personality and environment, and that the process of socializing the mentally deficient "may be described as that of so modifying or controlling the environmental stimuli in relation to the personality as to induce on the part of that personality the desired responses in behavior." (Pp. 358–359.) The final chapter sets forth a desired program for this group: (1) Specialized and industrial training from an early age so that the individuals may be fitted to take their places in the community; (2) special training in institutions and colonies for the more difficult cases with the hope of reconditioning them for return to the community; (3) organized community supervision to furnish guidance to all mentally deficient; (4) permanent segregation for those who cannot be socialized. In brief, a large majority of the mentally deficient may be safely and profitably retained as socially and economically functioning members of society provided they have the proper amount and kind of social control.

The volume meets a need long felt for a more thorough treatise on mental defectives. Coming from such an authority as Dr. Davies and embodying so many excellent references and citations makes it doubly valuable.

H. G. Duncan.
University of New Hampshire.

Exploring Religion with Eight Year Olds, by Helen Firman Sweet and Sophia Lyon Fahs. New York: Henry Holt and Co., 1930. Pp. IX + 238.

Character through Creative Experience, by William C. Bower. Chicago: The University of Chicago Press, 1930. Pp. XII + 276.

Every one who recalls his own Sunday School days with their lesson leaflets, golden-text cards, sunbeam songs, and their emphasis

on memorization which rendered glib and platitudinous some of the most inspiring literature of man will read Mrs. Sweet's book with a distinct sense of shock and more than a little delight.

"Exploring Religion with Eight Year Olds" is a moving account of what happened to one Sunday School class at the School of Religion of the Union Theological Seminary. Educators and psychologists will find it a sincere and well-done record of an experiment in democratic, child-centered education. Here are real children, quarreling, fighting, discussing everything from frogs to God, and gradually building up an attitude toward life and an increasing control of behavior. They wonder on what bases parents have the right to expect obedience. They consider what to do if some one hits them, and discard Robert's belief that to "sock him one" is best, in favor of the more rational approach of talking things over. They discuss the nature of God with almost scientific detachment, and they reject the "What would Jesus have me do?" of an older girl as a reliable guide to conduct! They go on an emotional spree over the sufferings of miners, but decide, a few Sundays later, that mine owners, too, have their problems.

The authors have done an excellent piece of work in presenting their material. Part One contains the human record of the exploration, the words, and acts of the children. Part Two interprets, gives the point of view, the standards by which growth was determined, and valuable suggestions for others who want to help children to find themselves. A well-selected bibliography is appended.

Every worker in progressive schools will profit by a reading of this book, not only for the vital picture of children which it gives, but also for the way in which the record is written. Few of the published records of the "new" schools are as clear and well balanced. Psychologists and research workers will value this record for the glimpse it gives of a world beyond and around their sometimes limited and naive dissections of isolated learning situations, the world in which real children live and move and have their being.

Finally, this work is significant because it points the way toward a vital, functioning-from-within kind of character education. If character development is ever to pass from the realm of lofty talk into that of actual achievement, experiments like this one must become everyday practice.

"Character through Creative Experience" examines the ultimate values of life against a background of the best scientific and philosophical techniques which the newer education has to offer. The author

is a professor of religious education at the University of Chicago. He maintains that the techniques of creative experience which have proved so effective in secular education are exactly those best adapted to accomplish the vastly more important and complex aim of character development.

He conceives education to be "the initiation of the young into a creative, personal, and social experience," and character to be the outcome of "inner, self-chosen, progressive realizations of values that are in the process of being transmuted into integrating, organizing, and motivating life purposes." He has limited his discussion to the intellectual factors of intelligence, reflective thinking, evaluation, and purposing, but insists that a many-sided approach is necessary. Within the limits set, it is an interesting and well-considered examination of the objectives of character education and the newer creative techniques based on self-choice. The chapters on "Education as Creative Experience," "The Unit of Learning," "The Discovery and Organization of Units," "The Steps of Self Learning," "Integrating Responses," and "Motivation" are particularly recommended. The book would be the better for a good bibliography.

ANN SHUMAKER.

Twins: Heredity and Environment, by Nathaniel D. Mttron Hirsch. Cambridge, Massachusetts: Harvard University Press, 1930. Pp. 158.

It has been remarked that the heredity-environment controversy is as pointless as many of the philosophical discussions of the ancients, and therefore should cease making demands on the time and attention of educators. But as long as men are interested in improving themselves and their children they will eagerly attempt to unravel these two powerful forces. Mr. Hirsch makes his contribution toward such unraveling in this little volume on twins.

Probably no subject in psychology is so seriously affected by emotional bias as the one treated in this book. The supporters of environment, steeped in the philosophy of democracy, make ludicrous claims for the all-powerful influence of training; the supporters of heredity go nearly as far in the other direction. As usual, the truth probably lies somewhere between. Facts about a truth seem to follow a normal probability curve; to be off too far in either the plus or minus direction is to tend to get away from the evidence. Hirsch does not so err.

His conclusion that "the contribution of heredity is several times as important as that of environment" is not the result of preconceived opinion, but grows out of the objective data which he has gathered.

The author's study of twins differs from other such studies in the method of selection used. As he points out, previous investigators have worked with unselected subjects, while he, on the contrary, chose selected subjects. "We intentionally selected," he says, "only those twins who were very similar, and those twins who were very dissimilar but of the same sex." He then proceeded to the selection of data on dissimilar twins living in similar environments, on similar twins living in similar environments, and on similar twins living in dissimilar environments. Concerning the first two approaches the writer makes the following concluding statement: "At least for the ninety-six pairs of twins in question . . . we can affirm that heredity is about five times as important as environment in respect to differences in intelligence quotient; about four times as important in respect to differences in head length; about four times as important in respect to differences in height; two and seven-tenths times as important in respect to differences in weight; about two times as important in respect to differences in cephalic index; and about one and one-half times as important in respect to differences in head width. Thus the relative importance of heredity versus environment in explaining differences varies very significantly in respect to the particular trait or form of ability that is measured."

The number of cases of separated similar twins is necessarily very small; hence, as Mr. Hirsch points out, the results are not statistically reliable. It is interesting, however, to note that "for four of the six comparisons the ratio of differences is almost zero." Heredity again.

The reviewer has no hesitation in highly recommending this book to students of the heredity-environment problem. It makes excellent use of what is probably the most fruitful and most dependable source of data on the relative contributions of inheritance and training; namely, the study of twins.

HERBERT A. CARROLL.
University of Minnesota.

An Introduction to Child Study, by Ruth Strang. New York: The Macmillan Co., 1930. Pp. 550. $2.75.

The author's difficulty in finding a non-technical, accurate, and readable book for persons interested in child study led her to publish

this one—"a panorama of childhood from birth through adolescence" which aims to serve both as a handbook for parents and parent-study clubs and as a text for an orientation course in child study, whether for advanced high-school or college students, nursery-school teachers, normal schools, or nurses' training schools.

Other books on child study, however excellent, have failed to meet all the needs of these individuals in ways which Strang avoids by organizing her material around stages of development rather than topics; by omitting technical or theoretical discussion; by treating the entire period of childhood; by supplying specific details and concrete illustrative situations; and by emphasizing the process of normal development and not problems and the problem child.

Strang divides the entire span of childhood into five periods: The neo-natal, the earlier and the later pre-school, the primary, the pre-adolescent, and the adolescent stages. Under each heading is given information concerning physical and mental development based on the best available sources; how and what the child at each stage learns, or should learn, is considered in some detail; and practical suggestions are given for teaching the child and for providing for him a suitable environment. A chapter on how to study the child is included for each developmental period except the last. The suggested daily schedule given for each period should prove particularly useful to parents, as well as the brief but serviceable discussions of the treatment of special problems arising during the various periods.

Full and annotated bibliographies and a list of questions and problems appended to each of the sections furnish suggestions for those interested in further study. Charts, illustrations, and a detailed index make the book exceedingly usable.

The comprehensiveness of this book has led to what may seem to those interested only in one phase of child development to constitute a deficiency. The parent or teacher of very young children, for example, might prefer a more thorough discussion of the individual developmental stages rather than the comprehensive but less detailed treatment of the entire span of childhood. But separate periods have been dealt with elsewhere, as the bibliographies given indicate. Justification for a comprehensive treatment such as this would seem to lie in the fact that the appreciation of any one period of childhood is enhanced by some understanding of preceding and later stages. Throughout the book discussions, even when briefly and pointedly given, are nowhere seriously inadequate. HAZEL E. CROCKETT.

Love in the Machine Age: A Psychological Study of the Transition from Patriarchal Society, by Floyd Dell. New York: Farrar and Rinehart, 1930.

The professed aim of this book is to make available, especially to parents, certain outlines of modern psychological theory restated in sociological and historical terms. It seeks to make explicit theories which heretofore have remained ill-defined in the minds even of experts in the field of social science. The theme is that patriarchal training does not fit children for living in our highly mechanized, modern Western civilization.

Leta S. Hollingworth's criteria of psychic adulthood were impossible under the aristocratic patriarchal régime. There the conception of landed property was fundamental to the then entirely military scheme of family life. To retain property in the hands of the family it was necessary to stunt psychosexual development of the children in such a way as to make them dependent upon their parents. The conflicts which consequently arose between biosexual maturity and psychosexual dependence were inadequately resolved in the practices of arranged marriage, homosexuality, prostitution, polite adultery, and sacred celibacy. The middle class, rising to power, strove to retain the sexual conventions of this society, the economic bases of which it had overthrown. This retention of outworn practices in a vastly changed society results today only in neuroses. To attain equanimity we need to face the facts squarely.

The development of patriarchalism, its conflict with modern needs, and the significance of the underlying psychological facts are dealt with by Mr. Dell shrewdly, understandingly, and with profound insight. Nor does he omit to warn those would-be moderns who, fleeing from patriarchal conventions, only rush into the same old snare disguised to look like new.

Tentative recommendations concerned with improved training involve the establishment of complete economic independence for adolescents, if necessary by means of unqualified subsidy; earlier marriages made possible by the use of the latest contraceptive devices; extra-domestic employment for both sexes; the postponement of having children until they are wanted; intensive and extensive training for and experimentation in matehood.

As an analysis of current phenomena this thesis contains nothing surprisingly new. It is in the identification and development of a basic theme in social history, and in the perhaps unnecessarily

lengthy yet nevertheless remarkably pertinent use of contemporary psychological and educational theories and practices, that lies its main claim to recognition. As an interesting and intelligent statement of some of the most recent findings concerning the individual and social implications of sex behavior "Love in the Machine Age" may be recommended as a valuable contribution. An excellent bibliography of some two thousand references is appended.

O. L. HARVEY.
School of Education, Harvard University.

Understanding the Adolescent Girl, by Grace L. Elliott. New York: Henry Holt and Co., 1930. Pp. 134. $1.25.

As a gentle guide to adolescence, full of wise comments and sound judgment based on experience with adolescent girls, this interestingly written little volume excellently represents the sort of book which educators would do well to encourage parents, counselors, and adolescent girls themselves to read.

Concisely yet fairly comprehensively it describes the adolescent girl in her personal-historical and social setting; outlines the major adjustments which she seeks to make; briefly discusses the problems which are most commonly met with by adolescent girls; offers to counselors sensible suggestions concerning the manner of helping girls in meeting their problems; and finally exposes to the adult advisor her own possible defects so that she may thus the more understandingly collaborate with adolescents in making their adjustments in changing social relationships.

The work does not claim to be an original contribution; it is essentially a tactful introduction. The short well-selected bibliographies appended to each chapter indicate the way to more advanced reading.

O. L. HARVEY.
School of Education, Harvard University.

Educational Biology, by John C. Johnson. New York: The Macmillan Co., 1930. Pp. XX + 360.

This is a book designed to give prospective teachers a foundation in biology, which, together with psychology and sociology, is considered to be basic for sound educational theory and practice. Educational biology is not a new kind of biology, but only a selection of those

aspects of the subject which are judged to be most useful to the teacher.

The book presents a wide array of subject matter, covering everything from parasites to human eugenics. To insure accuracy in this broad field the author has gone to the length of having each chapter read by a specialist. The difficulty is well-suited to college freshmen.

The reviewer found the book intensely interesting, and, remembering his own terrible first course in biology some twenty years ago, is frankly envious of the youths who are to receive their introduction to the subject in this attractive way. Especially interesting are the accounts of evolution, inheritance, and the biographical sketches of eminent biologists. It is recommended not only as a text for freshmen but also as supplementary reading for any student of psychology who feels the need for a general survey of the field of biology.

MELVIN RIGG.

Kenyon College.

DEC 21 1950